

UNIVERSIDADE FEDERAL DO PARANÁ

LEONARDO GOMES

RECONSTRUÇÃO 3D DE ACERVOS CULTURAIS USANDO CÂMERAS RGB-D:
SOLUÇÃO DE COMPROMISSO ENTRE PRECISÃO E TEMPO APLICADA AO
PROJETO ALEIJADINHO DIGITAL

CURITIBA PR

2016

LEONARDO GOMES

RECONSTRUÇÃO 3D DE ACERVOS CULTURAIS USANDO CÂMERAS RGB-D:
SOLUÇÃO DE COMPROMISSO ENTRE PRECISÃO E TEMPO APLICADA AO
PROJETO ALEIJADINHO DIGITAL

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Informática no Programa de Pós-Graduação em Informática, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Olga Regina Pereira Bellon.

Co-orientador: Luciano Silva.

CURITIBA PR

2016

Ficha Catalográfica

G633r

Gomes, Leonardo

Reconstrução 3D de acervos culturais usando câmeras RGB-D: solução de compromisso entre precisão e tempo aplicada ao projeto Aleijadinho digital. / Leonardo Gomes. -- Curitiba, 2016.
178f. : il. [algumas color.] ; 30 cm.

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-graduação em Informática, 2016

Orientador: Olga Regina Pereira Bellon -- Coorientador: Luciano Silva.

1. Processamento de imagens -- Reconstrução 3D. 2. Acervos culturais. I. Universidade Federal do Paraná. II. Bellon, Olga Regina Pereira. III. Silva, Luciano. IV. Título.

CDD: 006.693

Termo de aprovação



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor CIÊNCIAS EXATAS
Programa de Pós Graduação em INFORMÁTICA
Código CAPES: 40001016034P5

ATA DE SESSÃO PÚBLICA DE DEFESA DE TESE PARA A OBTENÇÃO DO GRAU DE DOUTOR EM INFORMÁTICA

No dia vinte e nove de Abril de dois mil e dezesseis às 15:00 horas, na sala AUDITORIO, Rua Cel. Francisco H. dos Santos, 100 - Centro Politécnico da UFPR, do Setor de CIÊNCIAS EXATAS da Universidade Federal do Paraná, foram instalados os trabalhos de arguição do doutorando **LEONARDO GOMES** para a Defesa Pública de sua Tese intitulada: **"Reconstrução 3D de acervos culturais usando câmeras RGB-D: Solução de compromisso entre precisão e tempo aplicada ao projeto Aleijadinho Digital"**. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Professores Doutores: LUCIANO SILVA (UFPR), JORGE VICENTE LOPES DA SILVA (CTI), OLGA REGINA PEREIRA BELLON (UFPR), PAULO FABIANO URNAU GOTARDO (UFPR), RODRIGO MINETTO (UTFPR), SÉRGIO SCHEER (UFPR). Dando início à sessão, a presidência passou a palavra ao discente, para que o mesmo expusesse seu trabalho aos presentes. Em seguida, a presidência passou a palavra a cada um dos Examinadores, para suas respectivas arguições. O aluno respondeu a cada um dos arguidores. A presidência retomou a palavra para suas considerações finais e, depois, solicitou que os presentes e o doutorando deixassem a sala. A Banca Examinadora, então, reuniu-se sigilosamente e, após a discussão de suas avaliações, decidiu-se pela **APROVAÇÃO** do aluno. O doutorando foi convidado a ingressar novamente na sala, bem como os demais assistentes, após o que a presidência fez a leitura do Parecer da Banca Examinadora. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, LUCIANO SILVA, lavrei a presente ata, que vai assinada por mim e pelos membros da Comissão Examinadora.

Curitiba, 29 de Abril de 2016.

Prof LUCIANO SILVA (UFPR)
(Presidente da Banca Examinadora)

Prof OLGA REGINA PEREIRA BELLON (UFPR)
(Coorientadora)

Prof PAULO FABIANO URNAU GOTARDO (UFPR)

Prof SÉRGIO SCHEER (UFPR)

Prof JORGE VICENTE LOPES DA SILVA (CTI)

Prof RODRIGO MINETTO (UTFPR)

Dedico este trabalho as pessoas fascinantes que cruzaram meu caminho. Meus pais e familiares pela segurança e oportunidades. Os amigos, especialmente o Hashi, sem os quais e com os quais se fica louco. Meus orientadores Luciano Silva e Olga Regina Pereira Bellon que foram meus pais da vida acadêmica nos mais de 10 anos de formidável orientação. A todos os colegas presentes e passados do grupo IMAGO, pelo inestimável companheirismo. A Rebecca, por fazer parte da minha vida e me permitir fazer parte da dela. Aos amiguinhos virtuais presentes nos jogos eletrônicos, sem os quais provavelmente jamais teria despertado ainda na minha infância o interesse pela fascinante ciência de transformar códigos em soluções de problemas.

Agradecimentos

Agradeço ao CAPES, CNPq, IPHAN, UNESCO, UFPR e ao Prof. Dr. Antônio Fernando Batista pelo apoio.

Resumo

Neste trabalho é proposto a utilização de câmeras RGB-D dentro do contexto de digitalização 3D de acervos culturais. O grupo de pesquisa no qual este trabalho foi desenvolvido está atuando no projeto Aleijadinho Digital. Este projeto consiste da digitalização do acervo deixado pelo artista barroco Aleijadinho. Uma das contribuições desta tese está na apresentação de um estudo comparativo sobre os projetos mais relevantes da área da preservação digital. As diversas tecnologias de geração 3D destes e outros projetos de preservação cultural também são discutidas e classificadas. Os scanners tradicionalmente utilizados nestes trabalhos possuem alta precisão, chegando na casa dos micrômetros. No entanto podem custar até centenas de milhares de dólares, necessitam de tripés, fonte própria de energia, treinamento para calibrar e operar o equipamento, espaço físico amplo e a aquisição pode durar alguns minutos para uma única captura. Enquanto os sensores RGB-D apesar da menor precisão realizam até 60 capturas por segundo, são simples de serem operados, necessitam apenas da conexão com um notebook e em geral custam até 100 dólares. Diversos trabalhos recentes apresentaram métodos que exploram o potencial destes sensores. No entanto, estas abordagens com frequência apresentam deformações quando aplicadas em reconstruções 3D completas, 360 graus, para diversos objetos. Neste trabalho é proposto uma metodologia completa de reconstrução 3D baseada em câmeras RGB-D. E para atenuar o efeito da deformação é apresentado um método de alinhamento baseado em pares que expande o grafo de conectividade. Isto é feito através de um método original para detecção automática de novos pares de malhas 3D com região de sobreposição. Esta medida permite que eventuais erros de alinhamento sejam distribuídos de forma mais homogênea entre um maior número de pares de alinhamento. Esta técnica de alinhamento é importante pois o acúmulo de erro é intensificado pela grande quantidade de imagens geradas por câmeras RGB-D. Os experimentos foram realizados utilizando os dados de 30 obras do projeto Aleijadinho Digital, realizado em parceria com a UNESCO. Com o auxílio de modelos criados com um sensor 3D de precisão submilimétrica específico para tarefa de reconstrução 3D, experimentos comparativos foram realizados contra outros métodos estado da arte. Os resultados apresentados foram favoráveis ao método proposto no que diz respeito a precisão geométrica.

Palavras-chave: kinect, alinhamento, reconstrução 3D, acervos culturais.

Abstract

This work proposes a 3D reconstruction method using RGB-D cameras within the context of 3D digital preservation of cultural heritage. Our research group is involved in Digital Aleijadinho project, which consists of the digitization of the artworks from this baroque artist. We present a comparative study of the most influential projects of the digital preservation area, their contributions and 3D generation technologies. Scanners traditionally used in these projects can achieve micrometer precision. However they also can cost up to hundreds of thousands of dollars, require tripods, own source of energy, training to calibrate and operate the equipment, ample space and the acquisition may take several minutes for a single capture. While the RGB-D sensors despite the lower precision they achieve up to 60 fps, they are easy to operate, requiring only the connection to a notebook and generally cost up to \$100. Several recent studies presented methods that exploit the potential of these sensors. However, these approaches often exhibit errors when applied to full 3D reconstruction of different objects, known as the loop closure problem. This work proposes a complete methodology of 3D reconstruction based on RGB-D cameras. And to mitigate the loop closure effect we present an pairwise alignment method that expands the graph connectivity. This is done through a new method that expands connections in a pairwise alignment system by automatically discovering new pairs of meshes with overlapping region. Then the alignment error is distributed more evenly over the aligned pairs avoiding the loop closure of full 3D reconstructions. Our alignment approach is necessary because the error accumulation is intensified by the large amount of data generated by RGB-D cameras. The experiments were performed using the data of 30 artworks of Digital Aleijadinho project, conducted in partnership with UNESCO. Our experiments show detailed 3D models and favorably compare with state-of-the-art methods found in literature.

Keywords: kinect, alignment, 3D reconstruction, cultural heritage.

Sumário

1	Introdução	1
2	Preservação de acervos culturais	5
2.1	Principais projetos	5
2.2	Sensores e métodos de aquisição 3D adotados na preservação de acervos culturais	7
3	Trabalhos relacionados	15
3.1	Métodos de reconstrução em tempo-real	15
3.2	Métodos de reconstrução offline	18
3.3	Tempo-real x Offline	20
4	Método proposto	23
4.1	Aquisição e Pré-processamento	24
4.2	Alinhamento aos pares	27
4.3	Expansão do grafo de alinhamento	31
4.4	Integração	34
5	Resultados Experimentais	37
5.1	Base de experimental	37
5.2	Experimentos	37
5.2.1	Modelos - Método proposto x Cui <i>et al.</i>	40
5.3	Discussão sobre o experimentos	56
5.4	Comparativo	56
5.4.1	SmartScan	57

5.4.2	Kinect	58
5.4.3	Intel Realsense F200	59
5.4.4	Scanner de tempo de percurso	60
6	Conclusão	63
6.1	Trabalhos Futuros	65
	Referências Bibliográficas	67

Lista de Figuras

1.1	Sequência de imagens apresentando as etapas do sistema proposto utilizando as imagens da escultura de Nossa senhora das Mercês, obra atribuída ao artista barroco "O Aleijadinho": A) Conjunto original de imagens de profundidade e coloridas; B) Superfície 3D resultado da combinação de cada conjunto de imagens de profundidade na etapa de pré-processamento; C) Alinhamento aos pares utilizando <i>Iterative Closest Points</i> , ICP; D) Alinhamento global com a técnica proposta; E) Integração volumétrica por Poisson; F) Modelo final.	2
2.1	Scanner 3D e câmera digital acoplados em um tripé.	13
4.1	Etapas do método proposto.	24
4.2	Intel Realsense e Kinect. A imagem apresenta os dois sensores e seus principais componentes.	24
4.3	Matsuo <i>et al.</i> [Matsuo et al., 2013]. A image apresenta o antes e depois do método de Matsuo que atenua ruído e deduz dados de regiões em que o sensor não capturou a profundidade (exemplificado no quadro vermelho).	26
4.4	A figura apresenta a renderização da superfície de uma escultura de madeira em forma de tábua. Na direita o método de Richardt <i>et al.</i> [Richardt et al., 2012] e na esquerda a mesma cena reconstruída com um scanner de precisão submilimétrica. Na imagem de Richardt a tábua aparenta um aspecto planificado e deformado no comparativo com o modelo do sensor de maior precisão.	26

4.5	A figura representa duas malhas 3D renderizadas em verde e vermelho da escultura de Nossa senhora das Mercês. Apesar da região do tronco e braço esquerdo estarem adequadamente alinhados a mão direita (destacada em azul) nas duas malhas ficou aproximadamente 1cm deslocada. Isso ocorreu pois do ponto de vista do sensor, enquanto todos os pontos da malha verde possuem valores de distância próximos entre si, na malha vermelha a região do tronco estava apenas 90cm e a região da mão estava 130cm de distância.	27
4.6	Representação do alinhamento aos pares.	28
4.7	A figura apresenta as etapas principais do algoritmo de alinhamento utilizado no método proposto.	29
4.8	Representação do acúmulo de erro causado por alinhamento sequencial e posterior distribuição do erro.	32
4.9	Na figura é apresentada na parte da esquerda duas malhas 3D renderizadas em verde e vermelho de Nossa senhora das Mercês. Na direita os seus respectivos <i>bounding boxes</i> com sua intersecção em verde.	34
4.10	Na figura é apresentada duas reconstruções 3D da escultura de Santo Hilário e Nossa senhora das Mercês. Em amarelo está o alinhamento sequencial por ICP com Pulli apenas e em azul após o alinhamento global com o método proposto.	34
4.11	Na Figura é apresentado um modelo antes e depois do processo de suavização da superfície pelo método de Yutaka <i>et al.</i> [Yutaka and Ohtake, 2003].	35
4.12	Na figura são apresentadas duas reconstruções 3D da escultura de Nossa senhora das Mercês. Na esquerda são todas as malhas 3D alinhadas em um mesmo sistema de coordenadas global. Na direita os mesmos dados após o processo de integração renderizado como um <i>wireframe</i>	36
5.1	A imagem apresenta o erro relativo ao EQM ao objeto Anjo Tocheiro. Visualmente os modelos são semelhantes.	41
5.2	A imagem apresenta o erro relativo ao EQM ao objeto Crucifixo. O modelo de Cui <i>et al.</i> apresenta o problema de regiões faltantes na parte esquerda.	41

5.3	A imagem apresenta o erro relativo ao EQM ao objeto Imagem de vestir. Visualmente os modelos são semelhantes.	42
5.4	A imagem apresenta o erro relativo ao EQM ao objeto São Joaquim. Um adorno prateado arredondado foi colocado na cabeça da escultura durante a aquisição com o Kinect, esse objeto não estava presente na aquisição com o SmartScan (por conta das propriedades reflexivas da peça), por isso ele aparece colorido, para o método proposto, Cui <i>et al.</i> no entanto enfrentou o problema de regiões faltantes nesta área do modelo.	42
5.5	A imagem apresenta o erro relativo ao EQM ao objeto Nossa senhora das Mercês. Cui <i>et al.</i> passou pelo problema de regiões duplicadas na parte traseira do braço esquerdo da escultura.	43
5.6	A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo	
1.	Visualmente os dois modelos são semelhantes.	43
5.7	A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo	
2.	Visualmente os dois modelos são semelhantes.	44
5.8	A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo	
3.	Visualmente os dois modelos são semelhantes.	44
5.9	A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo	
4.	Na parte esquerda traseira do modelo gerado por Cui <i>et al.</i> existiu o problema da região duplicada.	45
5.10	A imagem apresenta o erro relativo ao EQM ao objeto Pastor Ajoelhado. Os dois modelos são visualmente semelhantes, a diferença na base do modelo gerado por Cui <i>et al.</i> aconteceu pois este modelo é verticalmente mais alongado. . . .	45
5.11	A imagem apresenta o erro relativo ao EQM ao objeto Pescador. A diferença na base do modelo de Cui <i>et al.</i> aconteceu pela região não ser adequadamente segmentada pelo método.	46
5.12	A imagem apresenta o erro relativo ao EQM ao objeto Presépio Negro. Os dois modelos são visualmente semelhantes.	46

5.13	A imagem apresenta o erro relativo ao EQM ao objeto Santa Bárbara. Os dois modelos são visualmente semelhantes.	47
5.14	A imagem apresenta o erro relativo ao EQM ao objeto Santa Luzia. A região da coroa da escultura teve erro acentuado para os dois modelo por conta da região possuir refletância e gerar dados incorretos. O modelo gerado por Cui <i>et al.</i> também duplicou o braço direito da escultura.	47
5.15	A imagem apresenta o erro relativo ao EQM ao objeto Santana Mestra 1. Os dois modelos são visualmente semelhantes.	48
5.16	A imagem apresenta o erro relativo ao EQM ao objeto Santana Mestra 2. Os dois modelos são visualmente semelhantes.	48
5.17	A imagem apresenta o erro relativo ao EQM ao objeto Santo Atanásio. O modelo gerado por Cui <i>et al.</i> duplicou uma pequena região na parte traseira da escultura e falhou em reconstruir o arte direita dela.	49
5.18	A imagem apresenta o erro relativo ao EQM ao objeto Santo Hilário. O modelo gerado por Cui <i>et al.</i> duplicou uma pequena região na parte traseira da escultura e falhou em reconstruir o arte direita dela.	49
5.19	A imagem apresenta o erro relativo ao EQM ao objeto São Francisco de Paula 1. Os dois modelos são visualmente semelhantes.	50
5.20	A imagem apresenta o erro relativo ao EQM ao objeto São Francisco de Paula 2. Uma pequena região faltante na parte da cabeça da escultura do modelo gerado pelo método de Cui <i>et al.</i>	50
5.21	A imagem apresenta o erro relativo ao EQM ao objeto São Francisco de Paula 3. Região faltante no cajado da escultura pelo modelo gerado por Cui <i>et al.</i>	51
5.22	A imagem apresenta o erro relativo ao EQM ao objeto São João da Cruz. Erros ocorreram por conta das regiões reflexivas dos objetos.	51
5.23	A imagem apresenta o erro relativo ao EQM ao objeto São João Nepomuceno. A parte esquerda da face da escultura é uma região faltante no modelo gerado por Cui <i>et al.</i>	52

5.24	A imagem apresenta o erro relativo ao EQM ao objeto São Manoel. Os dois modelos são visualmente semelhantes.	52
5.25	A imagem apresenta o erro relativo ao EQM ao objeto São Pedro Nolasco 1. Os dois modelos são visualmente semelhantes.	53
5.26	A imagem apresenta o erro relativo ao EQM ao objeto São Pedro Nolasco 2. Os dois modelos são visualmente semelhantes.	53
5.27	A imagem apresenta o erro relativo ao EQM ao objeto São Raimundo Nonato 1. Pequenas regiões duplicadas na mão esquerda do modelo gerado por Cui <i>et al.</i> .	54
5.28	A imagem apresenta o erro relativo ao EQM ao objeto São Raimundo Nonato 2. Os modelos são visualmente semelhantes.	54
5.29	A imagem apresenta o erro relativo ao EQM ao objeto São Simão Stock. Problemas causados pela reflexividade do objeto no topo da cabeça da escultura.	55
5.30	A imagem apresenta o erro relativo ao EQM ao objeto Senhor dos Passos. No modelo gerado por Cui <i>et al.</i> , algumas regiões na base não foram segmentadas pelo método.	55
5.31	A foto apresenta o momento da aquisição pelo sensor SmartScan 3D - Breuckman. Ele está na altura de 3 metros com auxílio de um tripé. A escultura em madeira representa São Pedro Nolasco e está localizada na igreja de Nossa Senhora do Carmo em Ouro Preto - MG.	58
5.32	A imagem apresenta modelos 3D de uma escultura do compositor Mozart. A primeira imagem é o modelo gerado com o Intel Realsense F200 com textura utilizando o método proposto. A segunda imagem é a renderização do mesmo modelo sem textura. A terceira imagem é o mesmo objeto gerado utilizando o método proposto porém capturado utilizando o Microsoft Kinect.	59
5.33	A figura mostra fotos dos objetos que compõe a base de dados disponibilizada por Cui <i>et al.</i> [Cui et al., 2010].	61

Lista de Tabelas

2.1	Tabela classifica as tecnologias de aquisição segundo as características de Remondino <i>et al.</i> [Remondino and Rizzi, 2010]. +1 significa que a tecnologia é favoravelmente comparada com as demais tecnologias, 0 significa que ela é neutra, e -1 significa um comparativo desfavorável.	12
3.1	Avaliação de cada método em relação as características: TR(Tempo-real) - Apresenta resultados dinâmicos; Off(Offline) - Ao menos uma etapa executada após a captura finalizada; LE(Larga escala) - Capaz de reconstruir objetos do tamanho de edificações e ambientes internos; NR(Não-rígido) - Consegue lidar com superfícies maleáveis, como pessoas; Con(Contextual) - Adotam uma base pré-classificada; MS(Múltiplos sensores) - Combinam mais de um sensor. . . .	21
5.1	Tabela comparativa apresentando tempo de execução e o EQM entre os modelos gerados com o Kinect em relação ao <i>ground truth</i> gerado com um scanner de precisão submilimétrica.	39
5.2	Tabela comparativa apresentando o EQM entre os modelos de buda comparados com modelos de sua impressão 3D.	60
5.3	Tabela comparativa o EQM entre os modelos gerados com o método proposto e Cui <i>et al.</i> em relação ao modelo gerado com um scanner de precisão submilimétrica.	61

Lista de Acrônimos

3D	Três dimensões
EQM	Erro Quadrático médio
ICP	<i>Iterative Closest Points</i> - Método de alinhamento de malhas que liga pontos próximos iterativamente.
<i>RGB-D Camera</i>	<i>Red Green Blue - Depth Camera</i> - Sensores que gera imagens de cor e profundidade
<i>ToF Scanner</i>	<i>Time of Flight Scanner</i> - Scanner por tempo de percurso

Capítulo 1

Introdução

A reconstrução 3D e sua aplicação na preservação de acervos culturais tem recebido atenção considerável nas últimas décadas [Gomes et al., 2014]. O grupo de pesquisa IMAGO, do qual os autores deste trabalho fazem parte, atualmente está envolvido na digitalização de esculturas barrocas de Antônio Francisco Lisboa, "*O Aleijadinho*", obras consideradas patrimônio Mundial pela UNESCO. Estas esculturas do século XVIII esculpidas em madeira pintada estão perdendo a sua aparência original devido as propriedades naturais de sua matéria prima e por terem passado por consecutivos processos de restauração. Por estas razões, se torna importante realizar o registro da geometria e textura atuais destas esculturas, a fim de apoiar eventuais processos de restauração, armazenar e preservar sua aparência de forma fidedigna, auxiliar a divulgação digital através de museus virtuais, e assim por diante [Silva et al., 2013, Pamplona Segundo et al., 2012].

Notoriamente, o trabalho de geração de modelos 3D para preservação digital de obras de arte exige equipamentos de alto custo, da ordem de dezenas ou centenas de milhares de dólares, como o SmartSCAN 3D-HE da empresa Breuckmann. Além do custo também existe a demanda de treinamento por parte de seus operadores. Porém nos últimos dez anos o cenário se modificou com o surgimento e aprimoramento de sensores 3D, que geram dados em tempo-real e são simples de operar. Estes equipamentos permitiram reduzir o tempo de aquisição de algumas horas, até mesmo dias, para poucos minutos. Esta classe de sensores é chamada de Câmeras RGB-D (Red Green Blue Depth Cameras) e destacam-se as câmeras de tempo de

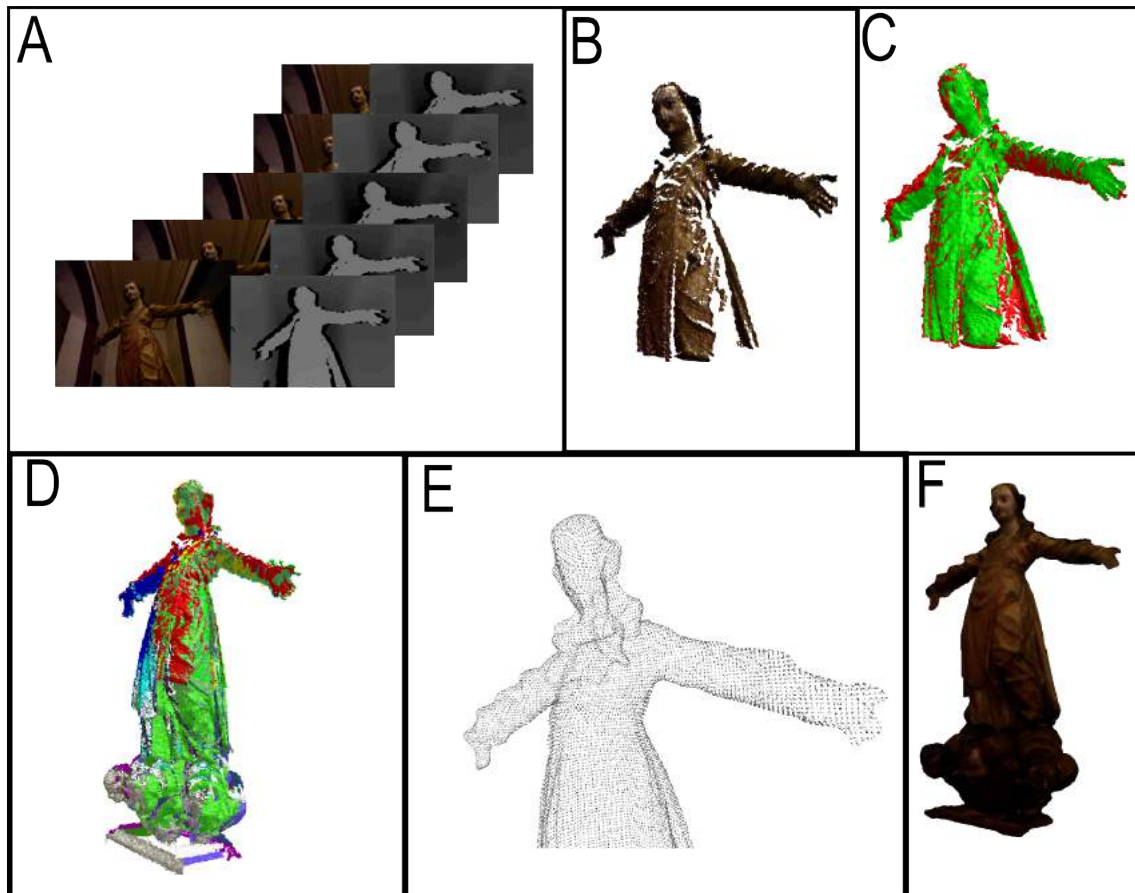


Figura 1.1: Sequência de imagens apresentando as etapas do sistema proposto utilizando as imagens da escultura de Nossa senhora das Mercês, obra atribuída ao artista barroco "O Aleijadinho": A) Conjunto original de imagens de profundidade e coloridas; B) Superfície 3D resultado da combinação de cada conjunto de imagens de profundidade na etapa de pré-processamento; C) Alinhamento aos pares utilizando *Iterative Closest Points*, ICP; D) Alinhamento global com a técnica proposta; E) Integração volumétrica por Poisson; F) Modelo final.

percurso [Callieri et al., 2009] e luz estruturada, como o Microsoft Kinect [Microsoft, 2014] e a linha Realsense da Intel [Intel, 2016].

Entre os métodos estado-da-arte para reconstrução 3D utilizando câmeras RGB-D destacam-se Cui *et al.* [Cui et al., 2013] pela precisão em seus modelos e *Kinect Fusion* [Izadi et al., 2011] pela precisão também obtida em tempo-real. O método aqui proposto apresenta resultados favoráveis no comparativo com o estado-da-arte além de possuir a importante característica da aquisição com *feedback* visual em tempo-real, embora exija um processamento posterior offline. Dessa forma é apresentada uma solução de compromisso entre velocidade de processamento e qualidade, priorizando a primeira em especial durante a aquisição e a segunda nas demais etapas.

A Figura 1.1 mostra uma sequência de imagens ilustrando as etapas do método desenvolvido. Inicialmente as imagens 3D são capturadas e salvas em disco. Se necessário um modelo parcial pode ser gerado para auxiliar no *feedback* visual da captura. Ao término desta etapa o modelo parcial é descartado, no caso de ter sido gerado, e os dados brutos salvos em disco passam pelo pré-processamento com aprimoramento dos dados. Na sequência, ocorre o alinhamento das imagens em um mesmo sistema de coordenadas para então serem integradas em um modelo final.

A maior contribuição do método proposto nesta tese está na solução original para expandir o grafo de ligamento entre malhas 3D na etapa de alinhamento, o que favorece a distribuição dos erros de cada par de malhas alinhadas, evitando o acúmulo que gera distorções visualmente evidentes no modelo 3D final. Este problema também é conhecido por *loop closure* [Pulli, 1999, Ho and Newman, 2007].

Este trabalho apresenta a seguinte **hipótese**: através do método proposto é possível realizar reconstruções 3D utilizando câmeras RGB-D com maior precisão do que o estado-da-arte até então. Este método viabiliza a utilização de câmeras RGB-D para diversas aplicações dentro do contexto de preservação digital de acervos culturais.

E para justificar essa hipótese temos os três seguintes **objetivos**:

1. Realizar uma revisão sobre os principais métodos e projetos de digitalização 3D de acervos culturais, em especial adotando câmeras RGB-D

2. Apresentar um comparativo qualitativo e quantitativo do método proposto contra o atual estado-da-arte
3. Apresentar uma discussão e comparação entre a qualidade dos sistemas que utilizam scanners de precisão submilimétrica e avaliar a viabilidade de utilizar câmeras RGB-D, de maior velocidade, em obras de arte.

Neste trabalho, são apresentadas soluções para reconstrução de modelos 3D, através de imagens de sensores RGB-D, com aplicação em cenários diversos. No entanto, a principal motivação é gerar soluções para preservação digital em 3D de acervos culturais. O trabalho se encaixa dentro de um projeto de pesquisa chamado, "Aleijadinho Digital", Andrade *et al.* [Andrade et al., 2012b], Santos *et al.* [Santos Junior et al., 2012] e Silva *et al.* [Silva et al., 2013].

Este trabalho está organizado da seguinte maneira: No Capítulo 2 é analisado alguns dos principais projetos de preservação digital e os diversos métodos aplicados; Já no Capítulo 3 são apresentados os trabalhos relacionados contendo os principais métodos tempo-real e *offline* utilizando câmeras RGB-D; O Capítulo 4 traz o método proposto e suas etapas destacando as contribuições e diferenciais para os demais métodos; No Capítulo 5 é mostrado os resultados experimentais comparados com outros métodos estado-da-arte e discussões sobre os mesmos; No Capítulo 6 são realizadas as considerações finais e sugestões de trabalhos futuros.

Capítulo 2

Preservação de acervos culturais

A preservação digital de acervos culturais, em especial através da digitalização 3D, é uma área cujo o interesse tem aumentado nas últimas duas décadas devido ao crescente avanço tecnológico. Entre as motivações principais são: (1) assegurar um registro fiel da forma e da aparência das obras, especialmente diante do risco de avaria por causas naturais ou acidentais ; (2) permitir a divulgação de coleções digitais para um grande público através de museus virtuais ; (3) criação de réplicas ; (4) permitir a identificação de falsificações ; (5) e para permitir fácil realização de avaliações específicas sobre a geometria ou textura quando estas forem inviáveis de obter a partir do objeto real.

Na sequência é apresentada alguns dos projetos mais influentes de digitalização de acervos culturais existentes na literatura.

2.1 Principais projetos

A partir do momento que foram desenvolvidos métodos viáveis de reconstrução 3D, os projetos de digitalização tridimensional de acervos culturais surgiram no mundo todo. Embora esta não seja uma lista completa, são discutidos alguns dos projetos pioneiros e mais importantes que desenvolveram suas sequência de métodos, também chamados de *pipelines*, que superaram dificuldades técnicas no processo de digitalização de acervos complexos.

O “**Projeto Michelangelo Digital**” por Levoy *et al.* [Levoy et al., 2000] descreve sua *pipeline* e a experiência da digitalização de 10 estátuas de grande porte criadas por Michelangelo,

dois interiores de edifícios, além de 1163 fragmentos de um mapa antigo feito em mármore na Itália. Eles adotaram scanners a laser de triangulação, scanners de tempo de percurso e câmeras digitais. Também são discutidos os desafios e soluções para trabalhar com esculturas de grande porte que chegam a gerar dois bilhões de polígonos para o maior objeto.

Outro trabalho importante é o **“Projeto o Grande Buda”** por Ikeuchi *et al.* [Ikeuchi et al., 2007] que descrevem o *pipeline* completo e os desafios para arquivar digitalmente e restaurar as três grandes estátuas de Buda localizadas no Japão medindo 2,7, 13 e 15 metros. O principal sensor adotado foi um scanner de tempo de percurso, ele emite um feixe laser e calcula o tempo que o feixe leva para refletir na superfície de cada região para estimar sua profundidade. Em seu artigo ele discute os desafios e soluções para lidar com um ambiente ao ar livre e um novo algoritmo para o registo e integração.

Além disso, há o **“Projeto Minerva”** por Fontana *et al.* [Fontana et al., 2002] que está relacionado com a Minerva de Arezzo, uma antiga estátua de bronze 1,55 metros de altura localizada na Itália. O modelo 3D é destinado a acompanhar as variações durante o processo de restauração da estátua. Um scanner 3D a laser de triangulação de alta precisão foi utilizado neste projeto. Trabalhando com um grande número de dados (cerca de um milhão de pontos em cada uma das 119 capturas) foi uma tarefa desafiadora dado o tamanho limitado da memória RAM e do poder de processamento dos computadores em 2001.

Existe também o **“Projeto Egito Eterno”** por Rushmeier [Rushmeier, 2006] que é um extenso projeto histórico que visa criar um guia digital e um museu virtual de artefatos do Egito antigo. Este projeto adota um sensor de tempo de percurso e um sistema de digitalização para imagens coloridas de alta qualidade Pro/3000 da *IBM Research*. Mais de 2000 varreduras 2D, 16 3D e quatro ambientes navegáveis compõe esta coleção virtual.

Outro trabalho é o **“Projeto Pieta”** por Bernardini *et al.* [Bernardini and Rushmeier, 2002, Bernardini et al., 2002] que foi responsável pela digitalização de uma grande escultura de mármore 2,26 metros de altura, localizada na Itália e criada por Michelangelo. Eles usaram um scanner de triangulação e uma câmera de vídeo para a textura. Entre as principais contribuições deste trabalho está o melhoramento na textura

obtida pela estimativa das propriedades intrínsecas da superfície calculadas por uma função de refletância de distribuição bidirecional (BRDF).

O **“Projeto Catedral Beauvais”** por Peter *et al.* [Allen et al., 2003] é um trabalho pioneiro de preservação de acervos culturais responsáveis pela digitalização da Catedral de Saint-Pierre em Beauvais, França. Foi utilizado um scanner de tempo de percurso laser (ToF scanner) e câmeras digitais para a aquisição. Entre as contribuições está um novo método para o alinhamento automático de imagens de profundidade.

Já o **“Projeto Templos Angkorian”** por Sonnemann *et al.* [Sonnemann et al., 2006] apresentou o trabalho de digitalização de uma grande área que contém mais de 1.000 estruturas históricas em Angkor, Camboja, usando imagens aéreas. É baseado em modelagem 3D fotogramétrica e mostra resultados de relevância para a documentação de paisagens arqueológicas e grandes patrimônios culturais.

E por último o **“Projeto Grande Buda de Bamiyan”** por Grün *et al.* [Grün et al., 2002] que apresenta os métodos utilizados na reconstrução 3D das duas grandes estátuas de Buda que medem 38 e 53 metros localizados em Bamiyan, no Afeganistão. Eles adotaram três técnicas de fotometria diferentes a partir de imagens de diversas câmeras para recuperar a geometria das obras destruídas pela milícia Talibã no ano de 2001.

2.2 Sensores e métodos de aquisição 3D adotados na preservação de acervos culturais

Nas últimas duas décadas, os novos sensores de profundidade têm contribuído para melhorar as informações geométricas e de cor dos modelos 3D para o campo da preservação digital. A variedade de cenários exige soluções diferentes que impactam diretamente sobre a escolha do sensor mais adequado. Para este fim, há características importantes a ter em conta, tal como apresentado por Remondino *et al.* [Remondino and Rizzi, 2010]:

- **Qualidade:** A precisão de um sensor afeta a fidelidade do modelo final.

- **Preço:** O custo dos sensores pode variar significativamente. Por exemplo: o Microsoft Kinect custa cerca de 40 dólares, enquanto um Smart Scan 3D-HE da Breuckmann custa dezenas de milhares de euros. Dado que o orçamento das missões de preservação digital pode ser limitado, esta é uma questão importante a considerar.
- **Portabilidade:** Restrições como falta de energia elétrica e dificuldade de acesso ao local são problemas comuns para certas campanhas de digitalização de acervos culturais. Portanto, é importante levar em consideração esses aspectos antes de optar por um sensor.
- **Tempo de aquisição:** Acervos culturais podem estar localizados em locais com um grande número de visitantes ou trabalhadores, entre outras restrições. Então é interessante que o equipamento minimize o tempo para aquisição dos dados gerando o mínimo de transtorno. O tempo de aquisição de um sensor pode também impactar no custo da campanha de digitalização por refletir no número de horas de trabalho e custos em estadias para a equipe.
- **Flexibilidade:** Acervos culturais têm dimensões, materiais e topologias arbitrárias. O sensor deve ser capaz de lidar com tais questões.

No entanto, até o presente momento, não há qualquer sensor que garanta todas essas características. Por conseguinte, uma solução possível pode ser a combinação de sensores diferentes para conseguir uma reconstrução 3D completa. Isto é particularmente importante para projetos que possuam esculturas complexas. Por exemplo, Bok *et al.* [Bok et al., 2011] apresentaram um sistema composto por quatro câmeras digitais e dois scanners a laser capazes de digitalizar grandes arquiteturas em um sítio cultural distribuídos por vários quilômetros quadrados de área.

Com o objetivo de obter maior fidelidade e realismo a qualidade é uma propriedade de muita importância para um sensor. De acordo com Curless *et al.* [Curless, 2000] a qualidade de uma medida tem quatro atributos:

- **Resolução:** Referente ao espaçamento entre os dados capturados.
- **Precisão:** Que se refere a variação estatística entre repetidas mensurações de um valor conhecido.

- **Repetibilidade:** Relacionada ao desvio estatístico do valor medido.
- **Sensibilidade ao ambiente:** Refere-se ao quão robustas são as medições diante de variações de luz, vento, temperatura, pressão, entre outras variáveis.

Blais [Blais, 2004], Godin *et al.* [Godin et al., 2002] e Pavlidis *et al.* [Pavlidis et al., 2007] apresentaram análises sobre os diferentes mecanismos de aquisição de dados 3D com ênfase nas aplicações de preservação de acervos culturais. A análise que segue é feita considerando os sensores mais comumente utilizados dentro de cada tecnologia. No entanto, é importante ressaltar que dois sensores distintos que adotam uma mesma tecnologia podem ter características muito variadas. Entre estes mecanismos de aquisição 3D são aqui listadas alguns dos mais relevantes. A Tabela 2.1 compara essas tecnologias entre si.

Triangulação à laser. O dispositivo emite um padrão de laser sobre o objeto e um sensor ótico, calibrado com o emissor de laser, identifica a posição desse padrão e calcula a informação de profundidade por simples triangulação. Devido à sua precisão, esses scanners são adotadas em campanhas de preservação digital, como Vrubel *et al.* [Vrubel et al., 2009], Arbace *et al.* [Arbace et al., 2012], Clara *et al.* [Baracchini et al., 2004] e Levoy *et al.* [Levoy et al., 2000]. Principais aspectos: precisão submilimétrica; refletância e transparência podem afetar as medições; alto custo (em geral os mais precisos estão na casa das dezenas de milhares de dólares).

Tempo de percurso. O sistema calcula o tempo de ida e volta dos impulsos de luz emitida pelo dispositivo para determinar a distância ao longo da superfície da cena. Callieri *et al.* [Callieri et al., 2009] propuseram um sistema composto por um scanner a laser de tempo de percurso suspenso sob um balão. Algoritmos para retificação que corrigem as distorções causadas pelo movimento do sensor foram aplicados durante a aquisição. O foco deste sistema foi principalmente a obtenção de dados de grandes prédios históricos vistos de cima. Principais aspectos: Ele opera através de grandes distâncias tornando-o adequado para a digitalização de grandes estruturas; dados ruidosos.

Luz estruturada. O dispositivo projeta um conjunto específico de padrões de luz e extrai a geometria das distorções destes padrões ao longo da superfície digitalizada. Os dispositivos nesta categoria têm diferentes tecnologias. Por exemplo câmeras RGB-D como a linhas Kinect da Microsoft e Realsense da Intel usam um único padrão de infravermelho e

pode adquirir dados 3D em uma taxa de até 30 imagens de profundidade por segundo com baixa precisão e dados ruidosos. O Smart Scan 3D-HE da Breuckmann utiliza um conjunto de luzes como padrão e pode adquirir imagens de profundidade com precisão submilimétrica. Podendo levar mais de 1 minuto para uma única imagem de profundidade [J. et al., 2004, Rocchini et al., 2001, Caspi et al., 1998, Guhring, 2001]. Principais aspectos: Sistema também captura textura; propriedades de superfície como refletância e transparência podem afetar as medições; geralmente sensíveis a iluminação ambiente.

Visão estéreo. O sistema é composto por dois sensores óticos calibrados que captura imagens simultaneamente. A distância entre os pontos correspondentes em ambas as imagens indicam o quão distante estão do ponto de vista do sistema e são utilizadas para calcular a medição de profundidade [Scharstein and Szeliski, 2003, Hertzmann and Seitz, 2005, Brown et al., 2003]. Principais aspectos: Baixo custo (webcams comuns podem ser adotadas); portabilidade; sistema também captura textura; baixa resolução e precisão (assumindo um ambiente não controlado); normalmente aquisição em tempo-real.

Forma a partir do movimento. O sistema gera a informação 3D a partir de uma sequência de imagens de um mesmo objeto, geralmente adquiridos utilizando uma câmera de vídeo. Pontos característicos são rastreados ao longo das imagens, a diferença na movimentação dos pontos baseia a criação de um espaço tridimensional que determina a posição relativa entre o objeto e a câmera. Existem também métodos que utilizam uma sequência de imagens e a informação do movimento relativo realizado entre objeto e sensor para resultados mais confiáveis [Szeliski and Kang, 1993, Tomasi and Kanade, 1992, Lakdawalla and Hertzmann, 2007]. Devido ao baixo custo e fácil aquisição, métodos baseados em imagens 2D também são populares, especialmente quando alta precisão não seja um requisito importante para a aplicação desejada, como mostrado por Snavely *et al.* [Snavely et al., 2006]. Principais aspectos: Baixo custo (webcams comuns podem ser adotadas); portabilidade; sistema também captura textura; baixa resolução e precisão (assumindo um ambiente não controlado).

Forma a partir da silhueta: Com uma coleção de imagens de um objeto com diferentes pontos de vista, a geometria do objeto é deduzida de sua silhueta. A informação de textura também é normalmente utilizada para aprimorar o processo. [Tosovic et al., 2002, Potmesil, 1987,

Laurentini, 1994]. Principais aspectos: Baixo custo (webcams comuns podem ser adotadas); portabilidade; sistema também captura textura; baixa resolução e precisão (assumindo um ambiente não controlado).

Forma a partir da sombra: Esta classe de métodos deduz a geometria de um objeto a partir de uma coleção de imagens sob iluminação variável. Também é possível combinar este método com os outros, a fim de melhorar a precisão [Zhang et al., 1999]. Principais aspectos: Baixo custo (webcams comuns podem ser adotadas); portabilidade; sistema também captura textura; baixa resolução e precisão (assumindo um ambiente não controlado).

Fotometria: Semelhante aos métodos de forma a partir da sombra, esta classe de métodos também deduz a geometria a partir de um conjunto de imagens de um objeto sob iluminação variável, mas assume um ambiente com fontes de luz e câmeras pré-calibrados. Com equipamento adequado em laboratórios especiais este método pode alcançar precisão semelhante aos scanners laser. [Basri and Jacobs, 2001, Hertzmann and Seitz, 2003]. Principais aspectos: precisão submilimétrica; precisa de um ambiente controlado.

Forma a partir do foco: O sistema requer uma coleção de imagens a partir do mesmo ponto de vista da captura com diferentes ajustes de foco. Por causa das lentes especiais necessárias, o método geralmente assume a utilização de um microscópio [Schechner and Kiryati, 1998]. Principais aspectos: exige um microscópio profissional; ele pode ser utilizado para pequenos objetos como insetos.

Digitalização por contato: Esta classe de método adota geralmente um braço robótico com variado grau de liberdade. O braço passa sobre a superfície do objeto tocando com a ponta de uma agulha que tem sua posição constantemente monitorada [Nashman et al., 1995]. Principais aspectos: precisão submilimétrica; aquisição lenta (pode levar várias horas para um único aquisição, dependendo do tamanho da superfície do objeto e taxa de amostragem desejada); exige contato com o objeto; o sistema é limitado as áreas acessíveis.

Topografia: O sistema é composto de uma estação total geodésica e é adequado para monumentos de grande escala. O equipamento é capaz de medir a distância e o ângulo de pontos característicos sobre o monumento. Estes pontos podem ser transformados em coordenadas

precisas [Livieratos, 1992]. Os principais aspectos: Aquisição lenta (dependendo da taxa de amostragem e o tamanho do objeto); é adotado para grandes superfícies.

Tecnologias	Qualidade	Preço	Portabilidade	Tempo de aquisição
Triangulação à laser	+1	-1	0	0
Tempo de percursso	-1	+1	+1	+1
Luz estruturada(RGB-D)	-1	+1	+1	+1
Visão estéreo	-1	0	+1	+1
Forma a partir do movimento	-1	+1	+1	+1
Forma a partir da silhueta	0	+1	+1	0
Forma a partir da sombra	0	+1	+1	0
Fotometria	+1	0	-1	0
Forma a partir do foco	+1	-1	-1	-1
Digitalização por contato	+1	0	-1	-1
Topografia	-1	-1	-1	-1

Tabela 2.1: Tabela classifica as tecnologias de aquisição segundo as características de Remondino *et al.* [Remondino and Rizzi, 2010]. +1 significa que a tecnologia é favoravelmente comparada com as demais tecnologias, 0 significa que ela é neutra, e -1 significa um comparativo desfavorável.

Sensores 3D óticos, como scanners a laser geram uma matriz 2D, chamada de imagem de profundidade, onde cada pixel guarda a distância de cada ponto na cena do ponto de vista do sensor. Com um processo de calibração adequado é possível transformar esses valores em informações 3D propriamente dita, tal como apresentado pela Besl [Besl, 1988]. É possível obter objetos integralmente digitalizados através da combinação destas imagens de profundidade sobrepostas a partir de diferentes pontos de vista. Sensores óticos normalmente não possuem câmeras coloridade com a mesma qualidade daquelas utilizadas por fotógrafos profissionais. No entanto, é possível utilizar uma câmara externa de alta resolução para capturar uma textura mais precisa, tal como apresentado por Iwashita et al. [Iwashita et al., 2005] e Andrade *et al.* [Andrade et al., 2012a]. Figura 2.1 apresenta um exemplo de um cenário de aquisição usando um scanner a laser 3D e uma câmara externa. Existem também alguns scanners que apresentam câmeras embutidas de alta resolução. Por exemplo, Mara *et al.* [Mara et al., 2010] geraram texturas, sem câmeras externas. Este projeto é para a digitalização 3D de tabuletas cuneiformes das coleções do Assyriology Heidelberg e Uruk-Warka Sammlung na Alemanha. Mas mesmo esses scanners podem se beneficiar de câmeras externas profissionais para geração de texturas

ainda mais precisas, tal como apresentado pela Akca *et al.* [Akca et al., 2007] e Bathow *et al.* [Bathow et al., 2010].



Figura 2.1: Scanner 3D e câmara digital acoplados em um tripé.

Métodos baseados em imagens consistem na detecção de propriedades como bordas, pontos de elevado gradiente, sombras entre outras características em um conjunto de imagens 2D para gerar um modelo 3D. Trabalhos importantes aplicados na preservação de acervos culturais, tais como aqueles apresentados neste capítulo, e que utilizam este tipo de abordagem são apresentadas por Snavely *et al.* [Snavely et al., 2006], Sinha *et al.* [Sinha et al., 2008] e Frahm *et al.* [Frahm et al., 2010].

Capítulo 3

Trabalhos relacionados

A popularização das câmeras RGB-D em especial o Microsoft Kinect motivou a pesquisa de reconstrução 3D para estes equipamentos. Diversos *Softwares* comerciais que adotam câmeras RGB-D para reconstrução 3D já estão disponíveis hoje, como por exemplo o *Skaneect* [Skaneect, 2014] e o *ReconstructMe* [ReconstructMe, 2014] que também oferece suporte para novos sensores como o Kinect 2 e os sensores Intel Realsense. Yi *et al.* [Yi et al., 2012] e Smisek *et al.* [Smisek et al., 2011] trazem uma visão global sobre a tecnologia de luz estruturada destes sensores e alguns métodos de reconstrução 3D para estes equipamentos.

Os trabalhos estão divididos em duas categorias: Os métodos exclusivamente em tempo-real, que adotam processamento paralelo em GPU; e os métodos que exigem ao menos uma etapa *offline*, que por não terem restrição quanto ao tempo de execução possuem mais opções quanto a técnicas de reconstrução geométrica.

3.1 Métodos de reconstrução em tempo-real

Atualmente, o método na literatura de maior relevância e citações para reconstrução 3D utilizando Kinect é o chamado *Kinect Fusion* [Izadi et al., 2011]. Este trabalho aplica uma combinação de estimativa de posicionamento de câmera baseado em uma variação própria do método *Iterative Closest Points*, ICP, de menor custo computacional para o alinhamento e representação volumétrica para integração das malhas capturadas. Os autores apresentaram

uma implementação fortemente baseada em paralelismo com GPU, tornando a metodologia de reconstrução em tempo-real.

Trabalhos posteriores ao *Kinect Fusion* surgiram para complementar limitações do método. O projeto *Kintinuous* [Whelan et al., 2012] propõe mudanças na estrutura de dados permitindo a digitalização de cenários de tamanho arbitrário. Whelan *et al.* [Whelan et al., 2013] apresentaram um algoritmo de alinhamento que traz deformações não-rígidas na malha gerada para atenuar acúmulo de erro de alinhamento. Chen *et al.* [Chen et al., 2013] adotam uma estrutura de dados hierárquica que dobra o desempenho sem perda de qualidade. Yue *et al.* [Yue et al., 2014] limitam a região de interesse otimizando a etapa de alinhamento e também dobrou o desempenho do método. Priscariu *et al.* [Prisacariu et al., 2014] apresentaram uma implementação *open source* do *Kinect Fusion* menos restritiva quanto ao hardware e mais eficiente no que diz respeito ao uso de memória e processamento de CPU e GPU. Pradeep *et al.* [Pradeep et al., 2013] apresentaram um método de reconstrução 3D em tempo-real inspirado no *Kinect Fusion*, porém utilizando somente uma webcam e atingindo resultados semelhantes em precisão.

Existem exemplos de trabalhos que buscam reconstrução 3D utilizando uma única imagem. Han *et al.* [Han et al., 2013] realizam reconstruções utilizando uma única imagem, porém refinando a profundidade utilizando modelagem de luz do ambiente e assumindo um objeto de refletância homogênea. Wu *et al.* [Wu et al., 2014] possuem uma metodologia que refina as imagens RGB-D baseado em um modelo de sombra com uma tecnologia em tempo-real que funciona em GPU. Zhang *et al.* [Zhang et al., 2014] apresentaram um método capaz de realizar reconstruções 3D utilizando somente uma imagem RGB-D combinando elementos da textura para refinar os dados de profundidade.

Muitos métodos trabalham com objetos não-rígidos como pessoas e tecidos. Wang *et al.* [Wang et al., 2014] apresentaram um método focado na reconstrução de faces. O rosto é dividido em regiões que passam por refinamentos sucessivos garantindo uma superfície suave mesmo diante de expressões faciais. Zollhofer *et al.* [Zollhofer et al., 2014a] apresentaram um sistema próprio composto por um par de câmeras que formam um sensor estéreo de baixo custo. Sua principal vantagem é que o método não faz nenhuma suposição sobre o objeto.

Também existem trabalhos que realizam reconstrução contextual vindo de uma base pré-classificada. Chen *et al.* [Chen et al., 2014] apresentaram um método de reconstrução 3D baseado em técnicas semânticas a partir de um sistema de relacionamento contextual aprendido a partir de uma base de dados com 10 mil objetos manualmente segmentados e categorizados. Este sistema aumenta a robustez do método para objetos conhecidos e também apresenta a vantagem do reconhecimento já estar embutido. Dame *et al.* [Dame et al., 2013] apresentam um método de reconstrução 3D em tempo-real utilizando Kinect que aplica informações semânticas, categorizando o item digitalizado em classes de objetos, para preencher regiões previamente sem informação de profundidade. O objeto é segmentado do plano de fundo por diferença de textura.

Alguns métodos são propostos para reconstruções em larga escala como interiores completos de casas e fachadas. Niebner *et al.* [Niebner et al., 2013] apresentaram um sistema de reconstrução 3D para larga escala que preserva detalhes utilizando uma estrutura de dados própria baseada em função hash, o método apresenta resultados semelhantes ao KinectFusion porém com significativo ganho em desempenho e gerência de memória. Steinbruecker *et al.* [Steinbruecker et al., 2013] mostraram um método para reconstrução 3D de multi-resolução para modelos considerados larga escala, tal como o interior de uma casa. Embora existam deformações em objetos menores a metodologia aplicada para alinhamento global é baseada em um grafo que armazena malhas 3D chave para evitar perda de referência no alinhamento. Jiang *et al.* [Jiang and Liu, 2013] propuseram um método para reconstrução 3D em tempo-real utilizando processamento em GPU. O método baseia-se na estratégia de dividir e conquistar para enfrentar o problema de larga escala sem perder precisão geométrica. Ren *et al.* [Ren et al., 2013b] apresentaram um método de reconstrução em tempo-real que adota informação de cor para auxiliar na etapa de alinhamento. Eles também propõe uma estrutura própria para gerenciar reconstruções em larga escala. Zollhofer *et al.* [Zollhofer et al., 2014b] aplicam uma técnica de reconstrução 3D de tempo-real em larga escala para preservação cultural de sítios arqueológicos na Turquia. Os modelos são utilizados para realizar medições e são aplicados para propósitos educacionais entre outras aplicações. O trabalho de Zollhofer *et al.* é um dos poucos existentes na literatura focados para digitalização de acervos culturais e que adota o Kinect. No entanto, diferente do trabalho aqui apresentado, eles aplicam seu método para fachadas e espaços abertos

em sítios arqueológicos e não para objetos digitalizados em 360 graus. Além dos trabalhos já citados existem outros de destaque na literatura para reconstrução 3D utilizando câmeras RGB-D em tempo-real. Kubacki *et al.* [Kubacki et al., 2012] apresentaram um pipeline completo para o alinhamento de malhas 3D, utilizando uma adaptação do ICP para alinhar progressivamente novas capturas diretamente na superfície representada por uma função de distância. Eles também apresentaram um método de integração de imagens de profundidade adaptado para paralelismo. Uma das limitações apontadas pelo autor no entanto é a representação por função de distância que pode perder alguns detalhes geométricos do objeto. O trabalho também possui a mesma limitação do *Kinect Fusion* quanto ao tratamento de *loop closure*. Ren *et al.* [Ren et al., 2013a] mostram um método de rastreamento e reconstrução 3D em um sistema probabilístico que reconstrói objetos a partir de primitivas geométricas simples e não necessita de coordenadas globais de alinhamento de múltiplas imagens. O método possui sua etapa de rastreamento em tempo-real.

O trabalho aqui proposto beneficia-se da aquisição em tempo-real aplicando diretamente uma implementação do *Kinect Fusion* para guiar a captura. Porém o restante do método não utiliza o modelo gerado nesta etapa na geração do modelo final, servindo apenas para auxiliar o operador durante a aquisição. Em relação ao *Kinect Fusion* o principal diferencial proposto é o tratamento do *loop closure*. Apesar do trabalho apresentado por Whelan *et al.* [Whelan et al., 2013] incluir um algoritmo para o tratamento deste problema, o método é voltado para o contexto de navegação robótica e não reconstrução de geometrias detalhadas. As deformações aplicadas para ajustar o modelo podem causar a perda de informação correta enquanto o método proposto reajusta e distribui o erro de alinhamento através de movimentos rígidos nas malhas capturadas.

3.2 Métodos de reconstrução offline

A literatura também apresenta métodos para reconstrução 3D a partir de câmeras RGB-D que não trabalham em tempo-real e necessitam ao menos uma etapa offline. Estes métodos podem utilizar uma maior variedade de técnicas por não possuírem a limitação com o tempo, além de serem compatíveis com computadores de menor poder de processamento.

Cui *et al.* [Cui et al., 2013] apresentaram os resultados mais precisos da literatura para câmeras RGB-D pelo que aponta seus experimentos. Eles utilizam super-resolução e filtros para atenuação de ruídos tanto randômicos quanto sistemáticos próprios de câmeras por tempo de percurso. Eles também utilizam técnica de alinhamento probabilístico para correção explícita do problema de *loop closure*.

A metodologia proposta neste trabalho se distingue do apresentado por Cui *et al.* [Cui et al., 2013] por não aplicar métodos de atenuação de ruídos e alinhamento específicos para as deformações causadas para um tipo de sensor (câmera por tempo de percurso). No método proposto não são feitas suposições sobre os dados e é aplicada uma metodologia de alinhamento própria para o tratamento do *loop closure* nas reconstruções 3D, evitando anomalias na geometria final.

Outros métodos se destacam para reconstrução offline, Chatterjee *et al.* [Chatterjee et al., 2012] apresentaram um método que adota um filtro bilateral adaptativo que aplica uma suavização mais intensa em regiões com superfícies mais homogêneas tornando as imagens menos ruidosas nessas áreas sem arriscar perder informação em regiões de maior curvatura. Noguchi *et al.* [Noguchi et al., 2013] combinam diversas imagens de profundidade e aplicam pesos locais para aumentar a resolução e aprimorar a precisão em seu processo de reconstrução 3D. O método foi desenvolvido visando imagens de profundidade com baixa resolução e com dados esparsos, tais como aqueles gerados pelo Kinect. Zhou *et al.* [Zhou and Koltun, 2013] apresentaram um método de reconstrução 3D utilizando câmeras RGB-D. Neste método alguns pontos de gradiente elevado são marcados como pontos de interesse que são armazenados para cada imagem e posteriormente utilizados na etapa de alinhamento global reduzindo o processamento. Bhavsar *et al.* [Bhavsar and Rajagopalan, 2012] apresentaram dois algoritmos. O primeiro é um método de super-resolução combinando imagens de profundidade próximas. O segundo para reconstrução contendo informação esparsa. Os dois algoritmos aplicam-se adequadamente ao tipo de dado ruidoso do Kinect.

Cappelletto *et al.* [Cappelletto et al., 2013] propõem uma metodologia completa utilizando Kinect que considera a informação de cor para auxiliar o alinhamento através de uma variação própria do ICP. Esta estratégia torna o método robusto para cenários em que a ausência

de detalhes geométricos dificulte a convergência do ICP, desde que o objeto possua detalhes de textura. No entanto, considerar a textura pode ser prejudicial em cenários onde a iluminação não seja controlada, pois a informação de cor pode apresentar-se de forma inconsistente ao longo da captura. Cappelletto *et al.* [Cappelletto et al., 2013] propõem a aplicação de uma solução para o *loop closure* semelhante ao método apresentado por Whelan *et al.* que é focada para as aplicações de navegação robótica.

Outra possibilidade é a combinação de diversos sensores que permitam cobrir áreas maiores e até mesmo aumentar a precisão dos dados ao custo de etapas extras de calibração. O trabalho de Alexiadis *et al.* [Alexiadis et al., 2013] apresenta um método para reconstrução 3D utilizando múltiplos Kinects em tempo-real para objetos maleáveis, incluindo pessoas. Nakazawa *et al.* [Nakazawa et al., 2012] também desenvolveram um método para reconstrução 3D completa e dinâmica em tempo-real utilizando um sistema composto por diversos Kinects pré-calibrados, porém não sincronizados entre si. Yilmaz *et al.* [Yilmaz and Karakus, 2013] e Somanath *et al.* [Somanath et al., 2013] também combinam sensores para reconstrução 3D. Estes métodos utilizam o Kinect em conjunto de um sensor passível de visão estéreo tornando a captura mais robusta. No entanto é necessário realizar uma etapa extra de calibração dos sensores.

Para a melhor visualização e efeitos de comparação a Tabela 3.1 lista algumas das principais características dos métodos discutidos.

3.3 Tempo-real x Offline

Algoritmos paralelizados em centenas de núcleos de GPU permitem otimização de métodos de reconstrução 3D de forma a atingirem resposta em tempo-real. As vantagens destes métodos são evidentes: maior velocidade do processo e *feedback* visual ao usuário durante a aquisição. Em uma primeira análise nos resultados é possível concluir que os métodos tempo-real estado-da-arte produzem modelos com qualidade equivalente aos métodos *offline*. No entanto, o efeito *loop closure* afeta a precisão dos métodos em temp-real.

Os métodos em tempo-real incrementam o modelo parcial em uma representação volumétrica a cada nova captura. E com o acúmulo de erros de alinhamento dessas capturas após

Trabalhos	TR	Off	LE	NR	Con	MS
<i>Kinect Fusion</i> [Izadi et al., 2011]	X					
<i>Kintinuous</i> [Whelan et al., 2012]	X		X			
Whelan <i>et al.</i> [Whelan et al., 2013]	X		X			
Chen <i>et al.</i> [Chen et al., 2013]	X		X			
Yue <i>et al.</i> [Yue et al., 2014]	X					
Priscariu <i>et al.</i> [Prisacariu et al., 2014]	X					
Pradeep <i>et al.</i> [Pradeep et al., 2013]	X					
Han <i>et al.</i> [Han et al., 2013]	X					
Wu <i>et al.</i> [Wu et al., 2014]	X					
Zhang <i>et al.</i> [Zhang et al., 2014]	X					
Wang <i>et al.</i> [Wang et al., 2014]	X			X		
Zollhofer <i>et al.</i> [Zollhofer et al., 2014a]	X			X		X
Chen <i>et al.</i> [Chen et al., 2014]	X				X	
Dame <i>et al.</i> [Dame et al., 2013]	X				X	
Niebner <i>et al.</i> [Niebner et al., 2013]	X		X			
Steinbruecker <i>et al.</i> [Steinbruecker et al., 2013]	X		X			
Jiang <i>et al.</i> [Jiang and Liu, 2013]	X		X			
Ren <i>et al.</i> [Ren et al., 2013b]	X		X			
Zollhofer <i>et al.</i> [Zollhofer et al., 2014b]	X		X			
Kubacki <i>et al.</i> [Kubacki et al., 2012]	X					
Ren <i>et al.</i> [Ren et al., 2013a]	X					
Cui <i>et al.</i> [Cui et al., 2013]		X				
Chatterjee <i>et al.</i> [Chatterjee et al., 2012]		X				
Noguchi <i>et al.</i> [Noguchi et al., 2013]		X				
Zhou <i>et al.</i> [Zhou and Koltun, 2013]		X	X			
Bhavsar <i>et al.</i> [Bhavsar and Rajagopalan, 2012]		X				
Cappelletto <i>et al.</i> [Cappelletto et al., 2013]		X				
Alexiadis <i>et al.</i> [Alexiadis et al., 2013]	X			X		X
Nakazawa <i>et al.</i> [Nakazawa et al., 2012]		X				X
Yilmaz <i>et al.</i> [Yilmaz and Karakus, 2013]		X	X			X
Somanath <i>et al.</i> [Somanath et al., 2013]		X				X

Tabela 3.1: Avaliação de cada método em relação as características: TR(Tempo-real) - Apresenta resultados dinâmicos; Off(Offline) - Ao menos uma etapa executada após a captura finalizada; LE(Larga escala) - Capaz de reconstruir objetos do tamanho de edificações e ambientes internos; NR(Não-rígido) - Consegue lidar com superfícies maleáveis, como pessoas; Con(Contextual) - Adotam uma base pré-classificada; MS(Multiplos sensores) - Combinam mais de um sensor.

uma trajetória longa, um desvio significativo pode ocorrer. Propostas de solução, como a feita por Whelan *et al.* [Whelan et al., 2013], envolvem estimar uma distorção do modelo até obter o encaixe adequado, porém o aspecto negativo dessa abordagem é não ter a garantia de que a distorção estimada corresponda ao objeto real.

Outra limitação dos métodos em tempo-real é a perda do referencial de alinhamento. Métodos de alinhamento não são livres de falhas e abordagens em tempo-real dependem de mecanismos de recuperação, que nem sempre são devidamente acionados e não garantem a recuperação deste referencial, sendo necessário por vezes recomençar o processo. Na Figura 4.10 o efeito desse acúmulo de erro de alinhamento é demonstrado.

Métodos offline não são necessariamente incrementais, como os métodos tempo-real, o que permite revisitar todas as capturas para gerar um alinhamento global adequado. Métodos offline também são mais robustos a falhas de alinhamento, visto que é possível efetuar correções manuais ou mesmo remover capturas incorretas que tenham sido afetadas pelo efeito visual de borramento na imagem.

Visando aplicações de preservação de acervos culturais, este trabalho opta por uma metodologia offline por permitir métodos robustos para alinhamento, implicando em geometrias mais realísticas. O método também conta com uma reconstrução em tempo-real para auxiliar a captura dos dados, facilitando a identificação da completude da aquisição.

Capítulo 4

Método proposto

No cenário de preservação digital de acervos culturais observa-se as seguintes características: objetos de tamanhos variados, restrição quanto a movimentação dos objetos, topologias desafiadoras com auto-oclusões, dentre outras limitações. A motivação do trabalho está em criar uma metodologia de reconstrução 3D utilizando câmeras RGB-D de baixo custo, capaz de lidar com as limitações destacadas e que tenha *feedback* visual em tempo-real e priorize a qualidade geométrica. Uma versão prévia do método proposto foi apresentado por Silva *et al.* [Silva et al., 2013]. Daquela versão para esta foi implementado um novo método para o pré-processamento de imagens, agora mais robusto, rápido e capaz de deduzir informações não capturadas através da sua vizinhança. E também foi gerado um novo método original de alinhamento global que leva em consideração especial o efeito *loop closure*. Sobre as implementações foram feitas em C++ e a principal biblioteca utilizada foi a OpenCV [Bradski, 2000] para operações básicas de carregar, salvar, acessar imagens. Os métodos são explicados em detalhes nas seções seguintes. A Figura 4.1 ilustra as etapas que serão apresentadas.

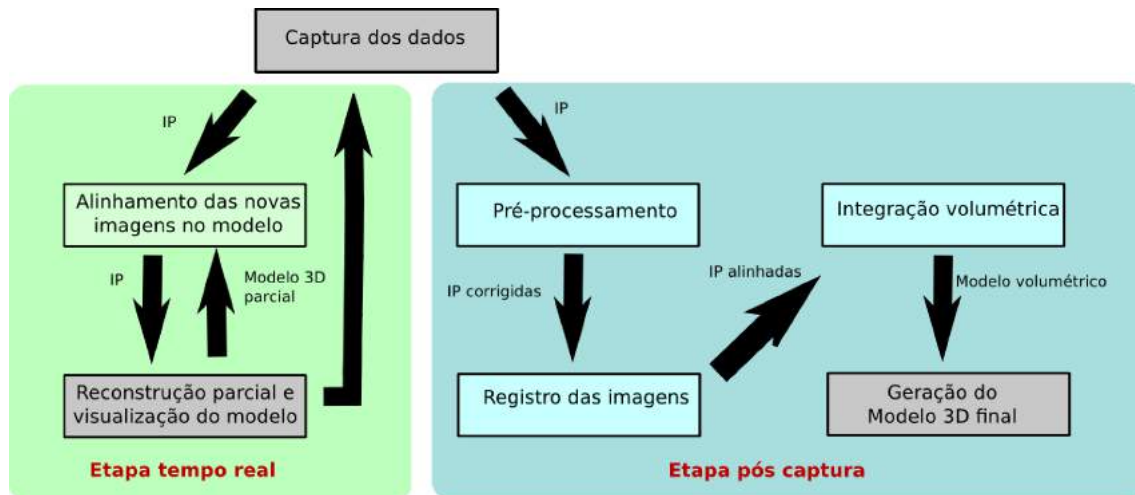


Figura 4.1: Etapas do método proposto.

4.1 Aquisição e Pré-processamento

Nos experimentos foram avaliados dois sensores de captura, o Microsoft Kinect e o Intel Realsense F200, ambos trabalham com a mesma tecnologia de luz estruturada e possuem características similares, abaixo será descrito o processo para o Kinect e ao final será comentado as diferenças para o sensor F200. Na figura 4.2 são apresentados os dois sensores.

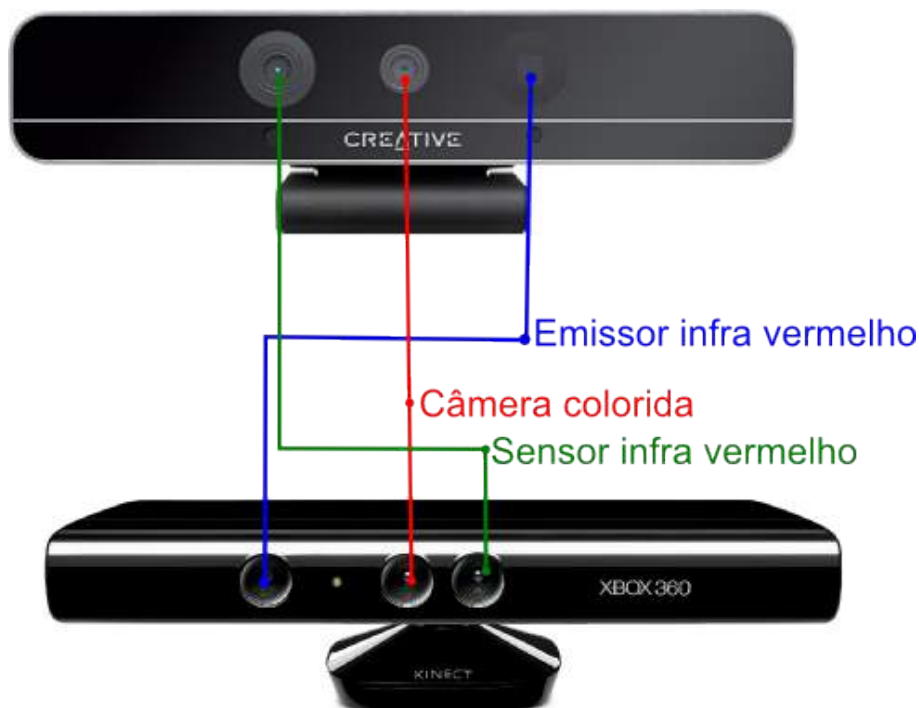


Figura 4.2: Intel Realsense e Kinect. A imagem apresenta os dois sensores e seus principais componentes.

A captura realizada com o Kinect foi feita de forma livre até cobrir toda a superfície do objeto, seja manipulando o sensor manualmente ou o objeto com o auxílio de uma base giratória. A completude da aquisição fica ao cargo do operador que é auxiliado por um modelo parcial. Este modelo parcial é reconstruído em tempo-real com o único propósito de que o operador possa acompanhar o progresso da digitalização, nada desse processo é utilizado para o modelo final. Esse modelo parcial é gerado utilizando o *Kinect Fusion* [Izadi et al., 2011]. Para o modelo final em si os dados originais são salvos em disco.

A etapa de captura produz N imagens de profundidade com resolução 640x480 acompanhados de N imagens coloridas de mesma resolução. As imagens de profundidade possuem erro médio de um milímetro quando capturados na distância de um metro do sensor. Uma análise completa sobre os dados do Kinect pode ser vista no artigo de Khoshelham *et al.* [Khoshelham and Elberink, 2012].

Para atenuar a baixa precisão dos dados é aplicado um método adaptado de Matsuo *et al.* [Matsuo et al., 2013]. Esta técnica trata-se de um filtro adaptativo que incrementa a qualidade da imagem de profundidade por atenuar ruídos e deduzir informação de profundidade de regiões onde os dados não puderam ser capturados. A Figura 4.3 demonstra o efeito deste método sobre uma imagem de profundidade. A região marcada em vermelho aponta uma região antes sem informação que através do método foi deduzida.

O método de Matsuo *et al.* [Matsuo et al., 2013] apresenta um filtro bilateral que substitui o valor de cada ponto de profundidade pela média ponderada por uma distribuição gaussiana de sua vizinhança porém considerando a imagem auxiliar de cor para calcular pesos pela diferença do ponto central da máscara. Um segundo filtro, semelhante ao da mediana, é aplicado visando remover borramentos que podem ser causados pelo primeiro filtro. Assim os pontos em regiões onde a diferença de cor de objetos em primeiro plano e a cor do plano de fundo sejam pequenas é recalculado.

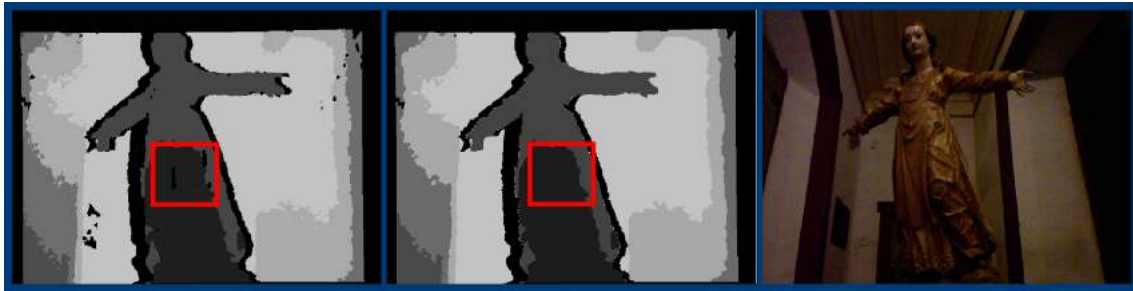


Figura 4.3: Matsuo *et al.* [Matsuo et al., 2013]. A image apresenta o antes e depois do método de Matsuo que atenua ruído e deduz dados de regiões em que o sensor não capturou a profundidade (exemplificado no quadro vermelho).

Além das informações da vizinhança temporal e espacial dos pontos é possível considerar outros aspectos como iluminação, sombreamento, refletância entre outras características. No entanto, tais características podem variar ao longo da captura quando se admite ambientes não controlados, o que gera deformações que inviabilizam a utilização destes dados em reconstruções 360 graus. Richardt *et al.* [Richardt et al., 2012] é um exemplo de método que considera todos estes elementos para refinar imagens de Kinect. Porém, o método sofre com as deformações citadas, como é apresentado na Figura 4.4. Para evitar esses efeitos, outras características não foram adotadas no processo de refinamento das imagens capturadas do método aqui proposto.

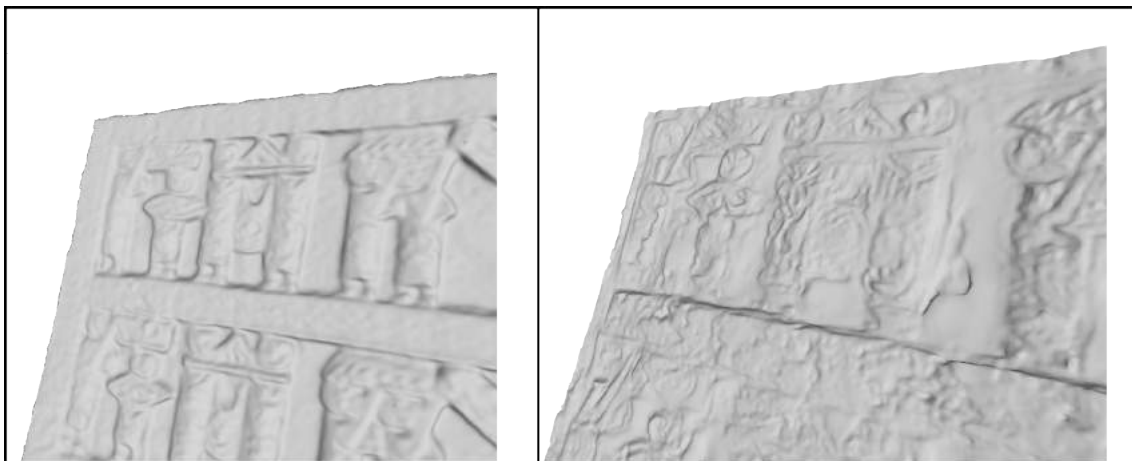


Figura 4.4: A figura apresenta a renderização da superfície de uma escultura de madeira em forma de tábua. Na direita o método de Richardt *et al.* [Richardt et al., 2012] e na esquerda a mesma cena reconstruída com um scanner de precisão submilimétrica. Na imagem de Richardt a tábua aparenta um aspecto planificado e deformado no comparativo com o modelo do sensor de maior precisão.

Mesmo com o sistema de câmeras e projetor bem calibrado o Kinect sofre com distorções, como é discutido no trabalho de Chow *et al.* [Chow and Lichti, 2013]. Esse efeito é

especialmente evidente quando em uma mesma captura o objeto possui algumas regiões próximas e outras distantes simultaneamente, conforme ilustrado na Figura 4.5. Uma maneira prática de reduzir este efeito e ao mesmo segmentar o objeto do fundo é considerar apenas pontos em uma faixa de profundidade, por exemplo pontos entre 80cm e 100cm de distância do sensor.

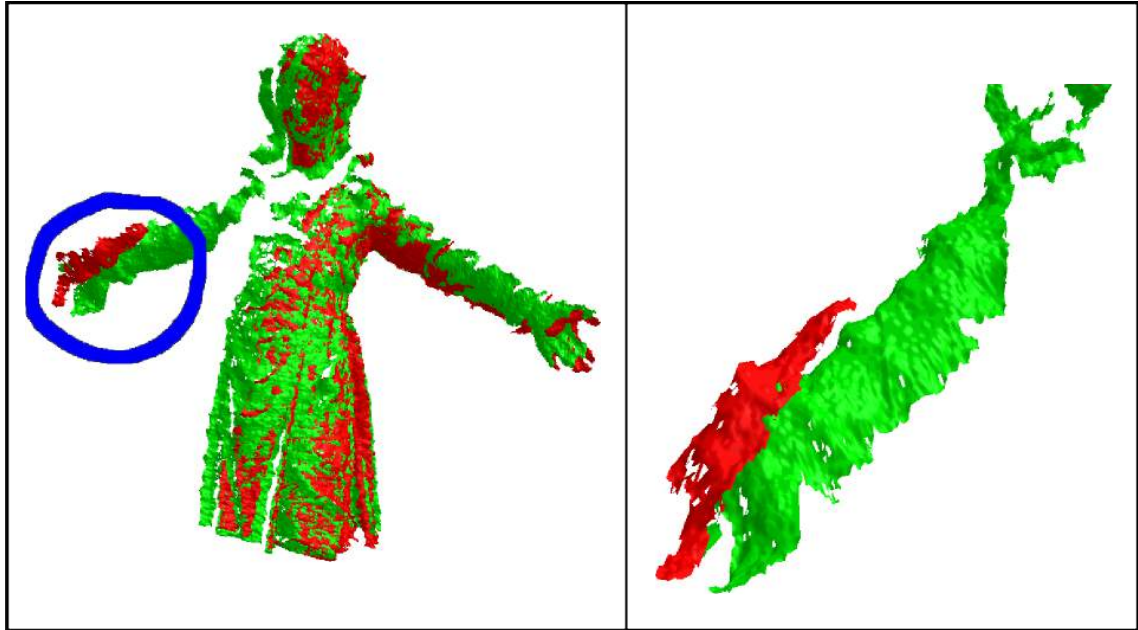


Figura 4.5: A figura representa duas malhas 3D renderizadas em verde e vermelho da escultura de Nossa senhora das Mercês. Apesar da região do tronco e braço esquerdo estarem adequadamente alinhados a mão direita (destacada em azul) nas duas malhas ficou aproximadamente 1cm deslocada. Isso ocorreu pois do ponto de vista do sensor, enquanto todos os pontos da malha verde possuem valores de distância próximos entre si, na malha vermelha a região do tronco estava apenas 90cm e a região da mão estava 130cm de distância.

Após este pré-processamento a saída consiste de imagens de profundidade segmentadas e refinadas pelo método de Matsuo. Dada a calibração do sensor as N imagens são transformadas em malhas 3D. Como as imagens estão em uma grade 2D a triangulação dos pontos é feita entre os vizinhos imediatos. Os triângulos perpendiculares ao sensor, ou seja, cuja normal tenha aproximadamente 90 graus de diferença para o ponto de vista do sensor são descartados pois não adequadamente visualizados em função da auto-occlusão.

4.2 Alinhamento aos pares

Depois do processo de aquisição, cada malha está no sistema de coordenadas local referente ao campo de visão do sensor no momento da captura. Portanto, torna-se necessário

colocar estes dados em um sistema de coordenadas global em uma etapa chamada de alinhamento, a Figura 4.6 ilustra esta etapa. Rusinkiewicz *et al.* [Rusinkiewicz and Levoy, 2001] e Salvi *et al.* [Salvi et al., 2007] apresentaram diversas técnicas para este fim, como por exemplo, o ICP que é adotado neste trabalho. Tam *et al.* [Tam et al., 2013] também apresentaram um survey contendo métodos de alinhamento, incluindo alguns que adotam o uso de Kinect. Dentro dos critérios deste trabalho o método aqui proposto é classificado como um método rígido baseado em restrições induzidas pela transformação. O Algoritmo 1 apresenta o método proposto, e a Figura 4.7 ilustra a sequência das etapas principais.

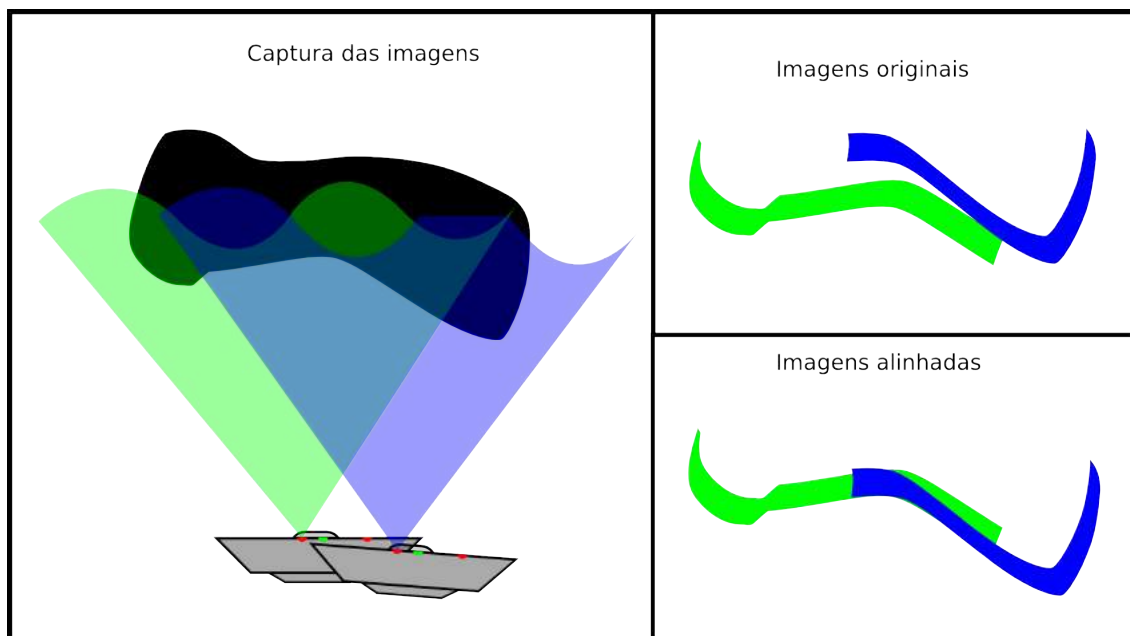


Figura 4.6: Representação do alinhamento aos pares.

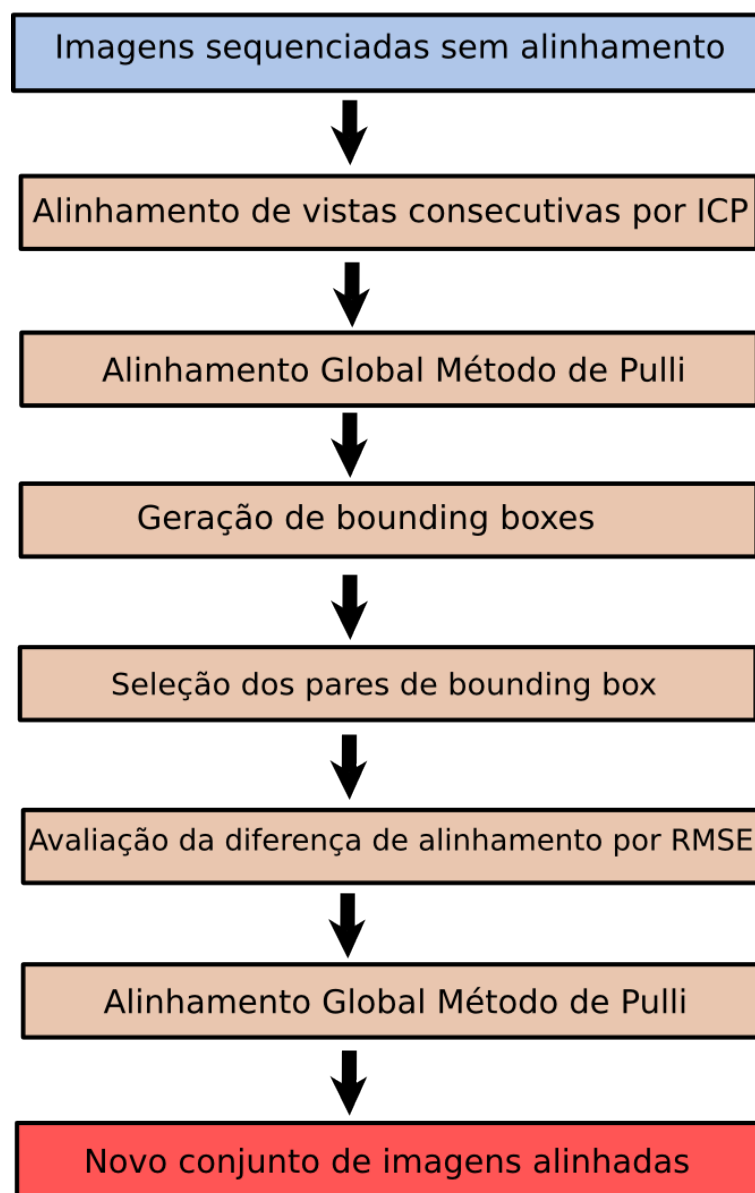


Figura 4.7: A figura apresenta as etapas principais do algoritmo de alinhamento utilizado no método proposto.

Algoritmo 1 Algoritmo de alinhamento proposto.

Requer: V composto por L malhas preprocessadas.

Garante: A composto por pares de alinhamento que alinham globalmente V

```

1: para todos  $vi \mid vi \in V$  faça
2:    $A = A + \text{ICP}(vi, vi+1);$ 
3: fim para
4:  $\text{pulli}(V, A);$ 
5: para todos  $(vi, vj) \mid vi, vj \in V ; vi \neq vj$  faça
6:    $\text{bbi} = \text{boundingbox}(vi);$ 
7:    $\text{bbj} = \text{boundingbox}(vj);$ 
8:   se  $\text{sobreposição}(\text{bbi}, \text{bbj})$  então
9:      $\text{erro} = \text{EQM\_ICP}(vi, vj);$ 
10:    se  $\text{erro} < \text{limiar1}$  então
11:       $A = A + \text{ICP}(vi, vj);$ 
12:    fim se
13:  fim se
14: fim para
15:  $\text{pulli}(V, A);$ 

```

Na etapa de aquisição a captura é realizada em até 30 imagens por segundo, gerando um volume de dados que pode somar dezenas de gigabytes, mesmo que para poucos segundos de captura. Devido ao tamanho dos dados, optou-se por trabalhar com alinhamento em pares, que não exige o carregamento de todos os dados simultaneamente em memória. Sendo assim, o primeiro desafio foi definir uma estratégia para gerar automaticamente os pares de alinhamento.

As imagens foram obtidas movimentando livremente a câmera, porém sem realizar movimentos bruscos durante a captura dos dados. Cada par de malhas consecutivas possuem um pequeno deslocamento, com mais de 95% de sobreposição nos casos avaliados, portanto a sequência original de captura nos garante um grafo conexo que interliga todas as malhas 3D (linhas 1,2 e 3 do algoritmo). Pelas mesmas razões o pré-alinhamento torna-se desnecessário nesta etapa, pois a diferença de posicionamento de duas malhas vizinhas assemelha-se ao

resultado esperado dos métodos de pré-alinhamento, o que aumentaria o custo computacional da abordagem proposta.

O alinhamento proposto é baseado em ICP, visto que suas variações são amplamente adotadas na literatura e apresentam resultados consistentes [Salvi et al., 2007]. Na abordagem proposta é utilizada como medida de erro o somatório do quadrado das distâncias ponto-a-plano de Chen e Medioni [Chen and Medioni, 1992]. Também foi utilizado *kd-tree* como estrutura de dados pois esta acelera significativamente a busca por pontos 3D próximos. O ICP foi dividido em duas fases, com parâmetros que priorizam primeiro a convergência seguida de outra fase com parâmetros que enfatizem uma maior precisão.

Na primeira fase do ICP foram utilizados somente um quarto dos pontos, melhorando o desempenho do processo. A distância máxima para correspondências foi de 9 mm. Este valor foi a melhor solução de compromisso para favorecer a convergência do ICP sem comprometer o desempenho. Na segunda fase todos os pontos foram adotados e apenas 1 mm de distância máxima de correspondência pois é equivalente ao erro de precisão do Kinect [Khoshelham and Elberink, 2012].

4.3 Expansão do grafo de alinhamento

É possível entender este problema de alinhamento como um grafo. Assumindo que exista um conjunto \mathbf{V} de malhas 3D que descrevem a superfície de um mesmo objeto. Existe também um conjunto \mathbf{A} formados por tuplas (a_i, a_j) para todo o a_i e a_j pertencentes a \mathbf{V} que possuem uma mesma subregião em comum, em outras palavras exista sobreposição entre as malhas 3D a_i e a_j . O grafo $\mathbf{G}(\mathbf{V}, \mathbf{A})$ deve ser sempre conexo, o que é garantido pela sequência de captura das câmeras RGB-D. Cada aresta possui um peso \mathbf{PS} , que representa a porcentagem da região de sobreposição dividido pelo erro de alinhamento gerado pelo ICP. (Ex.: Se a_j sobrepõe 0.8 da região de a_i , e erro de alinhamento médio é 2mm então $\mathbf{PS}(a_i, a_j) = 0.4$). O objetivo deste método é gerar um grafo \mathbf{G} conexo. Em que toda a tupla (a_i, a_j) cujo peso \mathbf{PS} seja inferior a um limiar¹ (esse valor depende do erro médio esperado para cada sensor) deve pertencer ao conjunto \mathbf{A} expandindo assim esse conjunto.

Para atingir este objetivo foi desenvolvido o método proposto para alinhamento global. Este método é auxiliado pelo algoritmo de Pulli [Pulli, 1999] (linha 4). Este algoritmo é amplamente adotado na literatura dada sua simplicidade e eficiência. Inicialmente declara-se uma malha como ativa e o restante como inativa. Na sequência os vizinhos inativos de uma malha ativa são adicionados a lista de ativos, um a um. A cada adição o erro de alinhamento é distribuído até que se atinja um equilíbrio. O algoritmo encerra quando todas as malhas tornaram-se ativas.

Essa etapa para diluir o erro de alinhamento é feita ao mover apenas uma malha por vez e alinhá-la com pontos virtuais. Os pontos virtuais não pertencem a nenhuma malha específica, mas são gerados dentro da intersecção entre as demais malhas. Este alinhamento é gerado utilizando a métrica ponto-a-ponto do ICP e mínimos quadrados.

O funcionamento do algoritmo de Pulli depende do grafo de conectividade das malhas 3D. Quanto maior o número de pares mais homogênea deverá ser a distribuição do erro de alinhamento, pois existirá um maior número de pontos virtuais. Como câmeras RGB-D produzem em média 30 aquisições por segundo e conseqüentemente pares de alinhamento sequenciais, o problema do acúmulo de erro é agravado. Por isso sozinho o método de Pulli não é capaz de evitar o *loop closure* para uma sequência linear composta por centenas de malhas 3D ruidosas ao exemplo do Kinect. A Figura 4.8 ajuda a explicar o efeito do *loop closure*.

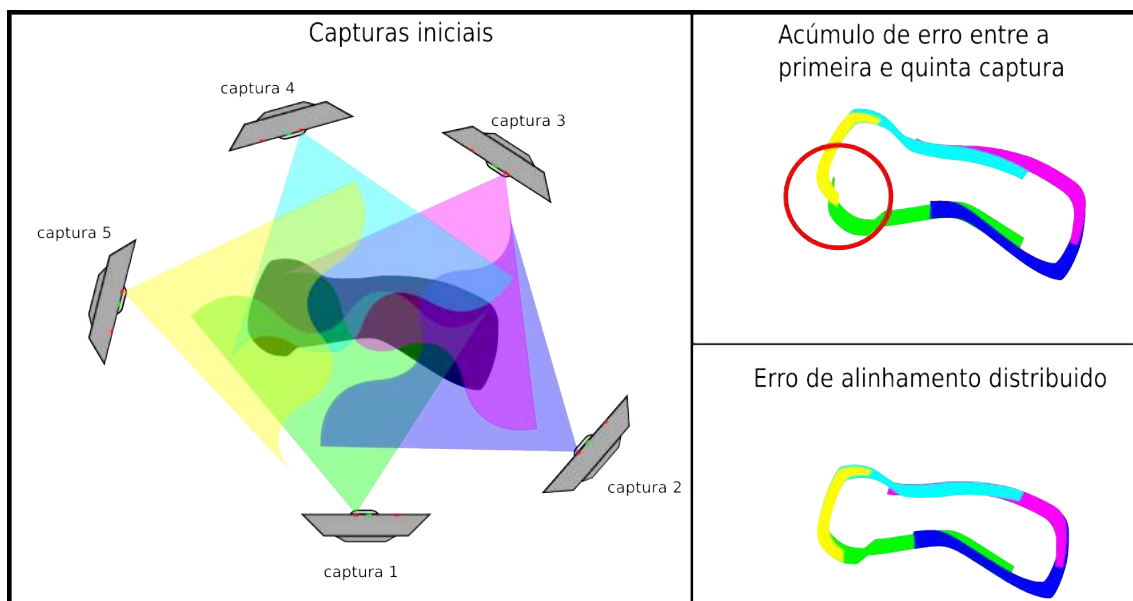


Figura 4.8: Representação do acúmulo de erro causado por alinhamento sequencial e posterior distribuição do erro.

Para solucionar a questão do *loop closure* foi desenvolvida uma abordagem original para aumentar o grafo de conectividade das imagens permitindo melhor desempenho para o método de Pulli. Para cada malha 3D das L geradas na etapa de pré-processamento é gerado o *bounding box*, ou seja, o paralelepípedo mínimo que engloba uma malha. O cálculo do *bounding box* é uma operação que requer apenas o armazenamento da informação das coordenadas mínimas e máximas de uma nuvem de pontos. Compara-se dois a dois todos os *bounding boxes* de cada malha em busca daqueles que possuam sobreposição (linhas 5, 6 e 7). Os pares que possuam sobreposição (linha 8) tem o seu alinhamento avaliado pela técnica do ICP previamente apresentada, caso o erro de alinhamento obtido pelo ICP seja inferior ao limiar1, relativo ao erro de precisão do equipamento (linhas 9 e 10), o par é adicionado ao grafo de conectividade. Finaliza-se o processo aplicando o algoritmo de alinhamento global de Pulli novamente (linha 15). Na Figura 4.9 são apresentadas duas malhas e seus respectivos *bounding boxes* com aproximadamente 40% de sobreposição.

Na Figura 4.10 é apresentado o modelo reconstruído de uma escultura de Nossa senhora das Mercês e de Santo Hilário com o alinhamento das malhas na sequência de aquisição junto do Pulli sem uso de métodos de expansão de grafo proposto. Observa-se que o acúmulo dos erros de alinhamento gerou deformações que descaracterizaram o objeto de seu modelo duplicando algumas regiões e simplesmente não completando outras.

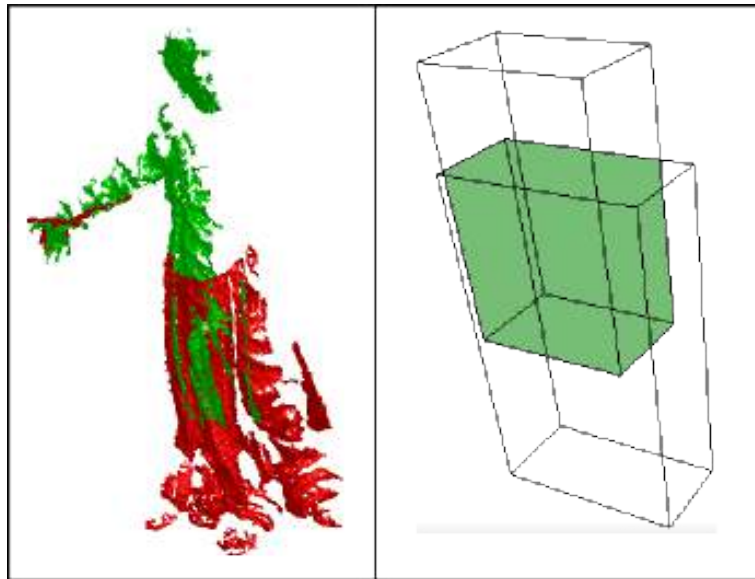


Figura 4.9: Na figura é apresentada na parte da esquerda duas malhas 3D renderizadas em verde e vermelho de Nossa senhora das Mercês. Na direita os seus respectivos *bounding boxes* com sua intersecção em verde.



Figura 4.10: Na figura é apresentada duas reconstruções 3D da escultura de Santo Hilário e Nossa senhora das Mercês. Em amarelo está o alinhamento sequencial por ICP com Pulli apenas e em azul após o alinhamento global com o método proposto.

4.4 Integração

Com todas as malhas 3D alinhadas em um mesmo sistema de coordenadas resta integrar os dados em um volume único. Foi adotado o método Poisson [Kazhdan et al., 2006] por

ser amplamente utilizado na literatura e ser especialmente robusto ao trabalhar com os dados ruidosos das câmeras RGB-D. O método IVIA [Santos Junior et al., 2012] também foi avaliado por apresentar em seus experimentos resultados de maior precisão no cenário de preservação digital ao comparar com Poisson. No entanto, não foi adotado pois este assume uma entrada de dados de alta resolução, sendo assim menos adequado as imagens do Kinect. Ao final para atenuar ainda mais os efeitos dos dados ruidosos, a malha final é suavizada através do método de Yutaka *et al.* [Yutaka and Ohtake, 2003] que considera pontos vizinhos cujas normais sejam aproximadamente paralelas, com isso a suavização ocorre minimizando a perda de detalhes geométricos, conforme apresentado na Figura 4.11. Já a Figura 4.12 apresenta antes e depois do processo de integração completo.

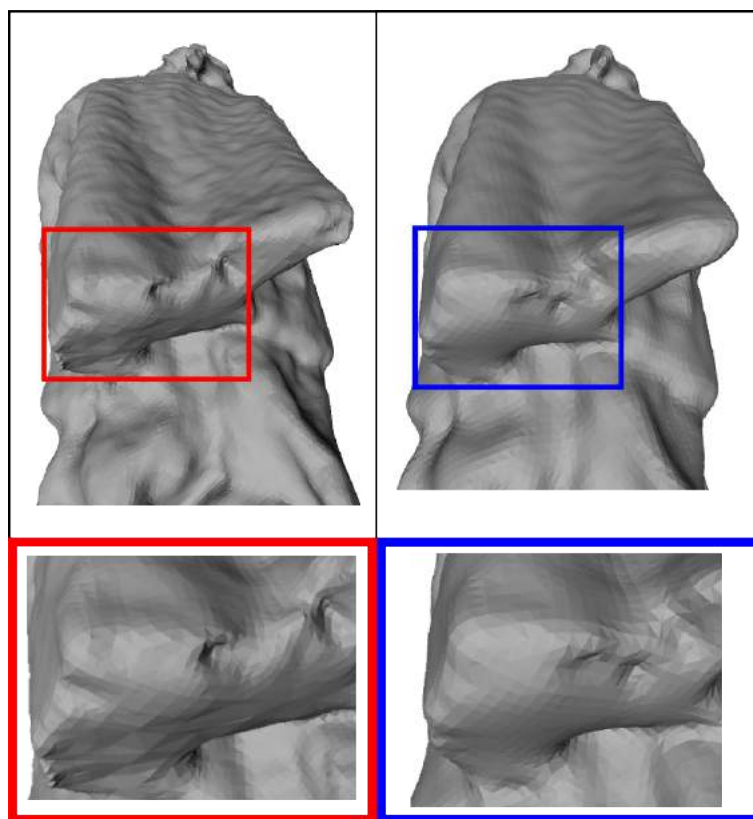


Figura 4.11: Na Figura é apresentado um modelo antes e depois do processo de suavização da superfície pelo método de Yutaka *et al.* [Yutaka and Ohtake, 2003].

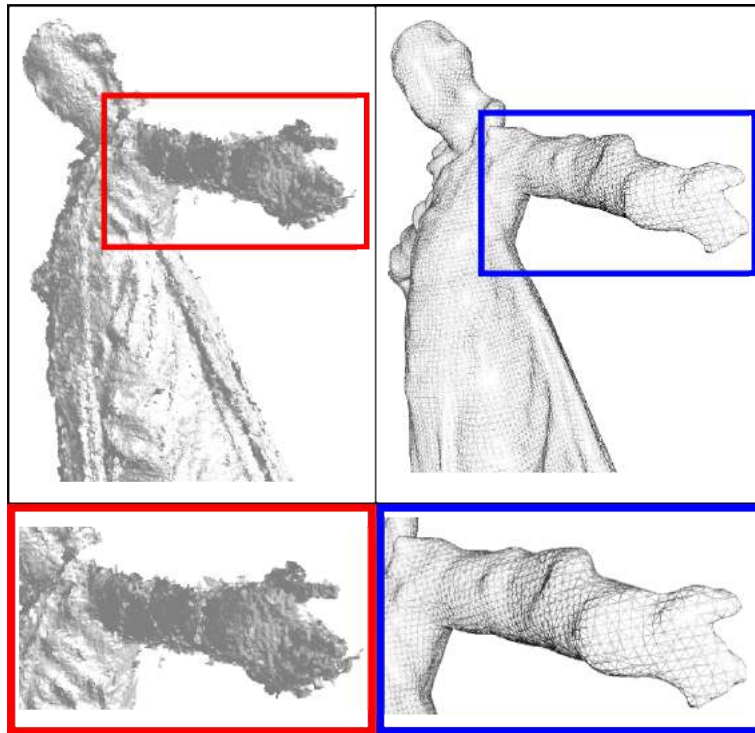


Figura 4.12: Na figura são apresentadas duas reconstruções 3D da escultura de Nossa senhora das Mercês. Na esquerda são todas as malhas 3D alinhadas em um mesmo sistema de coordenadas global. Na direita os mesmos dados após o processo de integração renderizado como um *wireframe*.

Capítulo 5

Resultados Experimentais

O método proposto foi avaliado na base de dados de um projeto de preservação de acervos culturais. A seguir esta base é apresentada em maiores detalhes, seguido dos experimentos em si e por fim a discussão sobre os resultados.

5.1 Base de experimental

Foram realizados testes utilizando 30 objetos de uma campanha real de digitalização de obras culturais em parceria com a UNESCO, projeto "Aleijadinho Digital". As obras estão dispostas em diversas igrejas e museus que são atribuídas ao artista barroco Aleijadinho na região da cidade de Ouro Preto - MG. As esculturas são talhadas em madeira pintada, que variam de 20 cm até 170 cm de altura. Com o Kinect a captura dessas imagens foi realizada na distância de aproximadamente um metro. Foram utilizados cerca de mil quadros cobrindo todas as superfícies dos objetos. Tanto o sensor quanto os objetos foram movimentados durante a aquisição. Além do Kinect as obras foram digitalizadas utilizando scanner de alta precisão modelo SmartSCAN 3D-HE da empresa Breuckmann.

5.2 Experimentos

A abordagem proposta foi comparada com outros dois métodos: O *Kinect Fusion*, pois junto de suas variantes é a metodologia tempo-real que apresentou resultados de maior

precisão; e Cui *et al.* [Cui et al., 2013] que realiza a reconstrução separado da aquisição em um processo posterior e apresentou os resultados mais precisos na literatura. Na subseção seguinte são apresentadas modelos reconstruídos a partir de dois objetos comparando suas geometrias obtidas através dos dois métodos estado-da-arte, o método proposto e da reconstrução por um scanner de alta precisão.

Foi realizado também uma avaliação quantitativa através do cálculo da raiz do quadrado das distâncias dos pontos mais próximos, erro quadrático médio (EQM), entre os modelos gerados com o Kinect comparado aos modelos gerados pelo scanner de maior precisão. Também foi calculado o tempo de execução para geração do modelo final pelos métodos avaliados e o resultado é apresentado na Tabela 5.1. Na avaliação de tempo foi utilizado um computador Intel(R) CPU 3.90GHz e 12 GB RAM.

Modelos	# malhas	KinectFusion		Cui <i>et al.</i>		Método proposto	
São Manoel	175	5.367631mm	tempo-real	6.856936mm	985s	4.551421mm	940s
Nossa Senhora da Mercês	344	7.842011mm	tempo-real	5.125663mm	1937s	4.214677mm	1848s
São Francisco de Paula 1	360	7.400198mm	tempo-real	7.250132mm	2027s	2.912708mm	1934s
Nossa Senhora do Carmo 1	162	9.357212mm	tempo-real	4.969834mm	912s	3.569082mm	870s
Santa Luzia	139	6.352581mm	tempo-real	5.814230mm	783s	2.950983mm	747s
São Joaquim	171	10.401252mm	tempo-real	3.547376mm	963s	3.296789mm	919s
Santo Hilário	163	6.532799mm	tempo-real	5.802813mm	917s	4.133509mm	876s
Santo Atanásio	194	8.410376mm	tempo-real	6.674652mm	1092s	5.884672mm	1042s
São João Nepomuceno	193	9.736920mm	tempo-real	3.127850mm	1087s	2.249081mm	1037s
Senhor dos Passos	121	5.683305mm	tempo-real	7.852160mm	681s	5.440989mm	650s
São Simão Stock	325	9.298310mm	tempo-real	3.586836mm	1830s	3.421987mm	1746s
Santana Mestra 1	130	7.956172mm	tempo-real	6.952797mm	732s	4.839201mm	698s
Anjo Tocheiro	421	7.238691mm	tempo-real	7.456862mm	2370s	5.920173mm	2262s
Pastor Ajoelhado	168	5.425417mm	tempo-real	4.234179mm	946s	2.842391mm	903s
Pescador	221	7.850283mm	tempo-real	6.193646mm	1244s	5.484328mm	1187s
Imagem de Vestir	356	9.307415mm	tempo-real	3.186762mm	2004s	3.188379mm	1913s
Presépio Negro	204	8.352581mm	tempo-real	7.305941mm	1149s	5.891273mm	1096s
Nossa Senhora do Carmo 2	185	10.831910mm	tempo-real	5.699038mm	1042s	3.409326mm	994s
Nossa Senhora do Carmo 3	186	7.194928mm	tempo-real	4.659273mm	1048s	4.918974mm	1000s
São Francisco de Paula 2	154	5.320231mm	tempo-real	8.352785mm	867s	4.308756mm	827s
Crucifixo	282	6.986201mm	tempo-real	4.959012mm	1588s	3.819874mm	1515s
São Raimundo Nonato 1	113	10.843595mm	tempo-real	4.821300mm	636s	3.308367mm	607s
São Pedro Nolasco 1	209	5.340816mm	tempo-real	3.454820mm	1177s	4.953482mm	1123s
São Raimundo Nonato 2	213	12.356712mm	tempo-real	5.897059mm	1199s	3.975349mm	1144s
São Pedro Nolasco 2	266	13.484158mm	tempo-real	6.764208mm	1498s	5.912908mm	1429s
São João da Cruz	271	9.183264mm	tempo-real	7.794641mm	1526s	2.208743mm	1456s
Nossa Senhora do Carmo 4	135	6.583482mm	tempo-real	4.148651mm	760s	2.019873mm	725s
Santana Mestra 2	121	10.674512mm	tempo-real	5.953292mm	681s	5.184324mm	650s
São Francisco de Paula 3	231	8.799548mm	tempo-real	8.342156mm	1301s	6.710984mm	1241s
Santa Bárbara	186	5.710629mm	tempo-real	3.345278 mm	1047s	3.859873mm	999s
Média	213	8.194104mm	tempo-real	5.671006mm	1201s	4.179415mm	1146s

Tabela 5.1: Tabela comparativa apresentando tempo de execução e o EQM entre os modelos gerados com o Kinect em relação ao *ground truth* gerado com um scanner de precisão submilimétrica.

Os resultados da Tabela 5.1 indicam resultados favoráveis em termos EQM para o método proposto em 27 dos 30 objetos avaliados. O tempo de execução foi superior ao método de Cui *et al.* e inferior ao tempo-real do *Kinect Fusion* para todos os objetos. Na próxima seção

os modelos gerados pelo Cui *et al.* e o método proposto são apresentados, na sequência tanto a tabela quanto os modelos são discutidos em detalhes.

5.2.1 Modelos - Método proposto x Cui *et al.*

As Figuras 5.1 até 5.30 ilustram o EQM ao renderizar regiões azuladas para erros menores e avermelhadas para erros maiores. O comparativo foi feito com o modelo gerado pelo scanner de precisão submilimétrica. Na parte superior dessas imagens é apresentado renderizações do método proposto e na parte inferior gerado com o método de Cui *et al.*, no canto inferior direito a foto do objeto e no canto superior direito a renderização colorida do scanner de alta precisão.

Os problemas principais encontrados no método de Cui *et al.* foram três:

1. **Regiões duplicadas.** Ocorrem quando dois agrupamentos de *meshes* que representam uma mesma região, não estão alinhados entre si. O erro de alinhamento acumulado aumenta a distância entre os dois grupos. Pode ser corrigido criando pares de alinhamento entre os dois agrupamentos.
2. **Regiões faltantes.** Semelhante ao caso anterior, ocorrem quando dois agrupamentos de *meshes* que representam uma mesma região, não estão alinhados entre si. Porém neste caso e o erro de alinhamento acumulado aproxima os dois grupo ao ponto que se atravessam, o método de integração não consegue resolver o volume para aquela região em que as normais dos *meshes* apontam para lados opostos. Pode ser corrigido criando pares de alinhamento entre os dois agrupamentos.
3. **Regiões reflexivas.** Certos objetos metálicos presentes nas esculturas refletem a luz e confundem os sensores, um problema comum para a totalidade dos métodos de reconstrução óticos.

Ao longo das imagens essas três questões e outras serão apontadas.

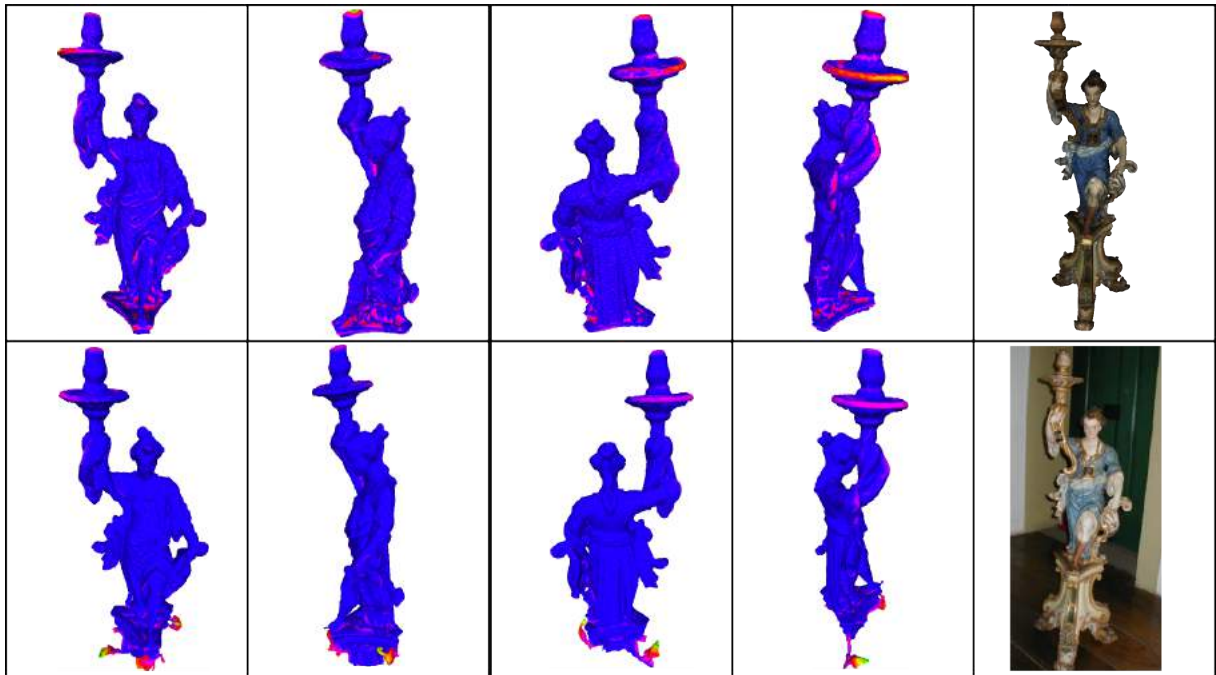


Figura 5.1: A imagem apresenta o erro relativo ao EQM ao objeto Anjo Tocheiro. Visualmente os modelos são semelhantes.

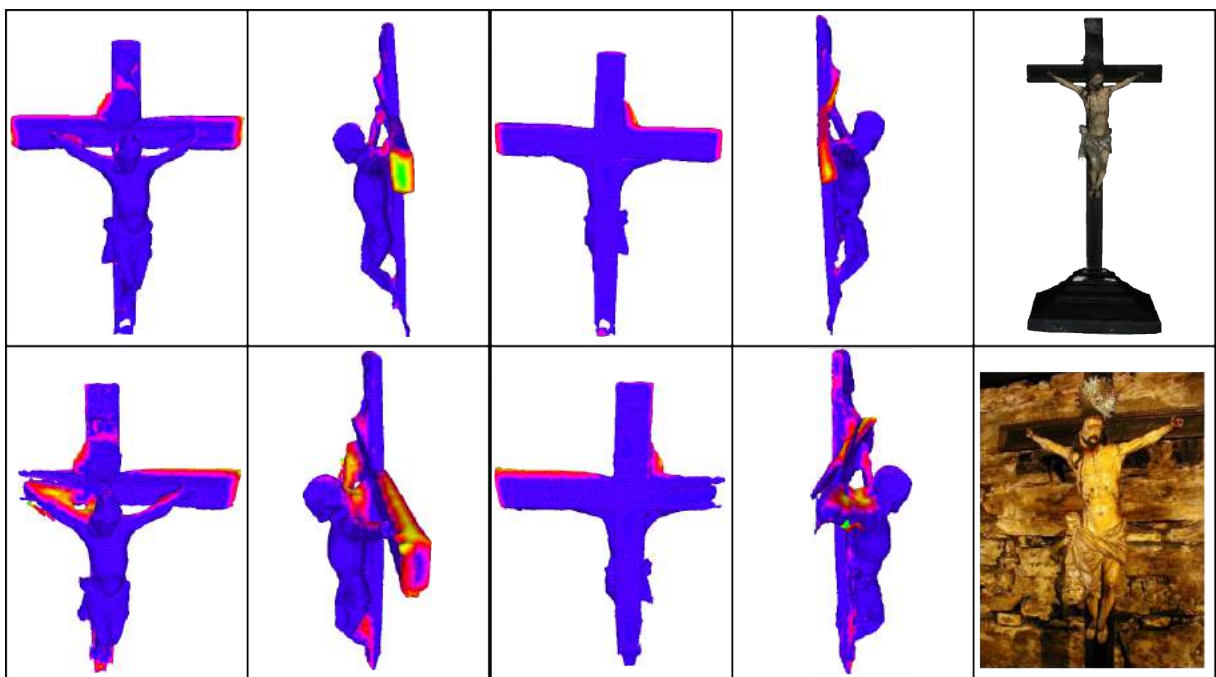


Figura 5.2: A imagem apresenta o erro relativo ao EQM ao objeto Crucifixo. O modelo de Cui *et al.* apresenta o problema de regiões faltantes na parte esquerda.

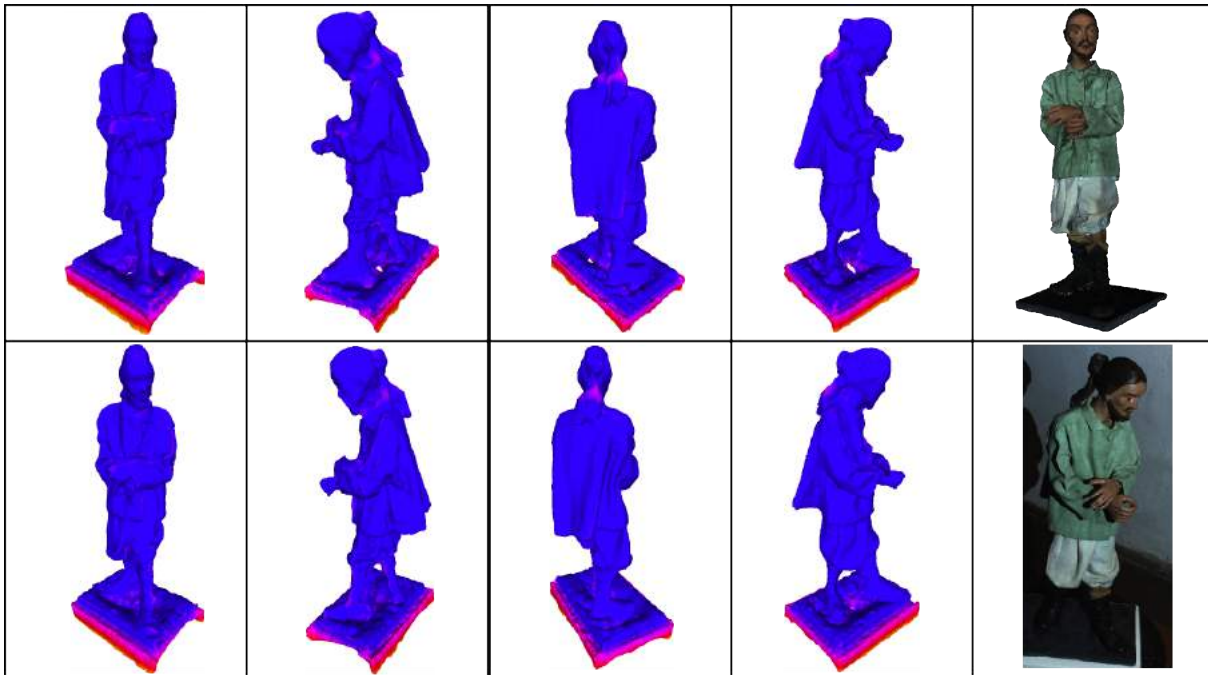


Figura 5.3: A imagem apresenta o erro relativo ao EQM ao objeto Imagem de vestir. Visualmente os modelos são semelhantes.

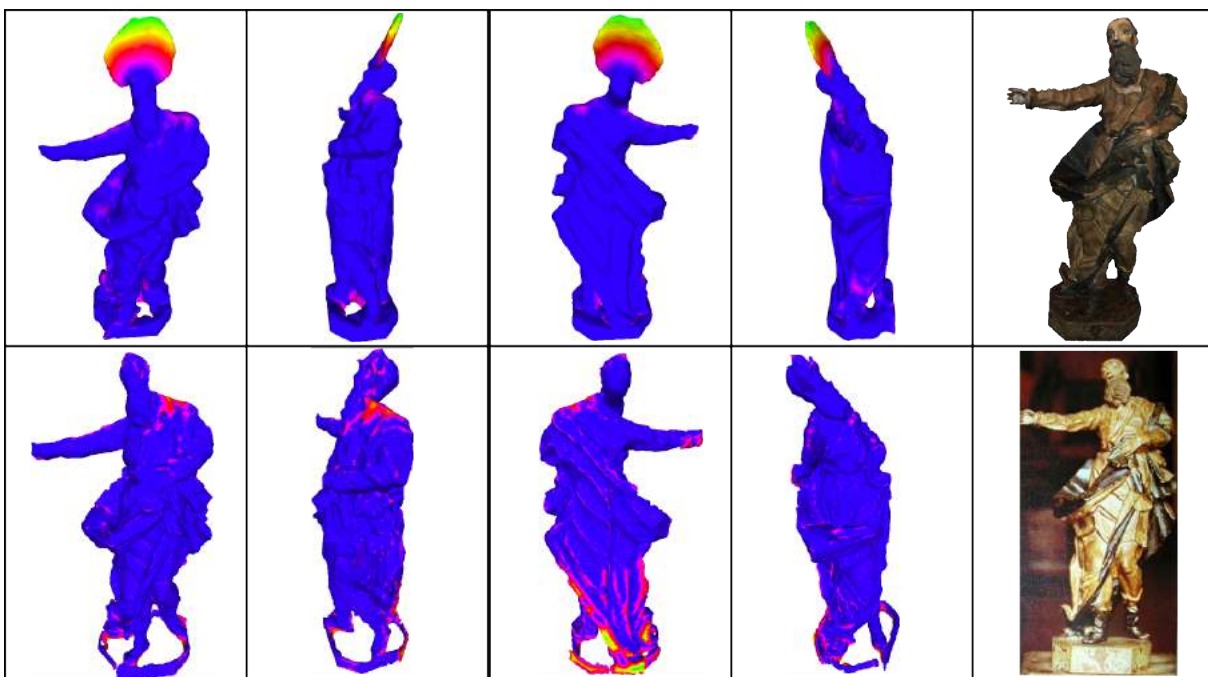


Figura 5.4: A imagem apresenta o erro relativo ao EQM ao objeto São Joaquim. Um adorno prateado arredondado foi colocado na cabeça da escultura durante a aquisição com o Kinect, esse objeto não estava presente na aquisição com o SmartScan (por conta das propriedades reflexivas da peça), por isso ele aparece colorido, para o método proposto, Cui *et al.* no entanto enfrentou o problema de regiões faltantes nesta área do modelo.

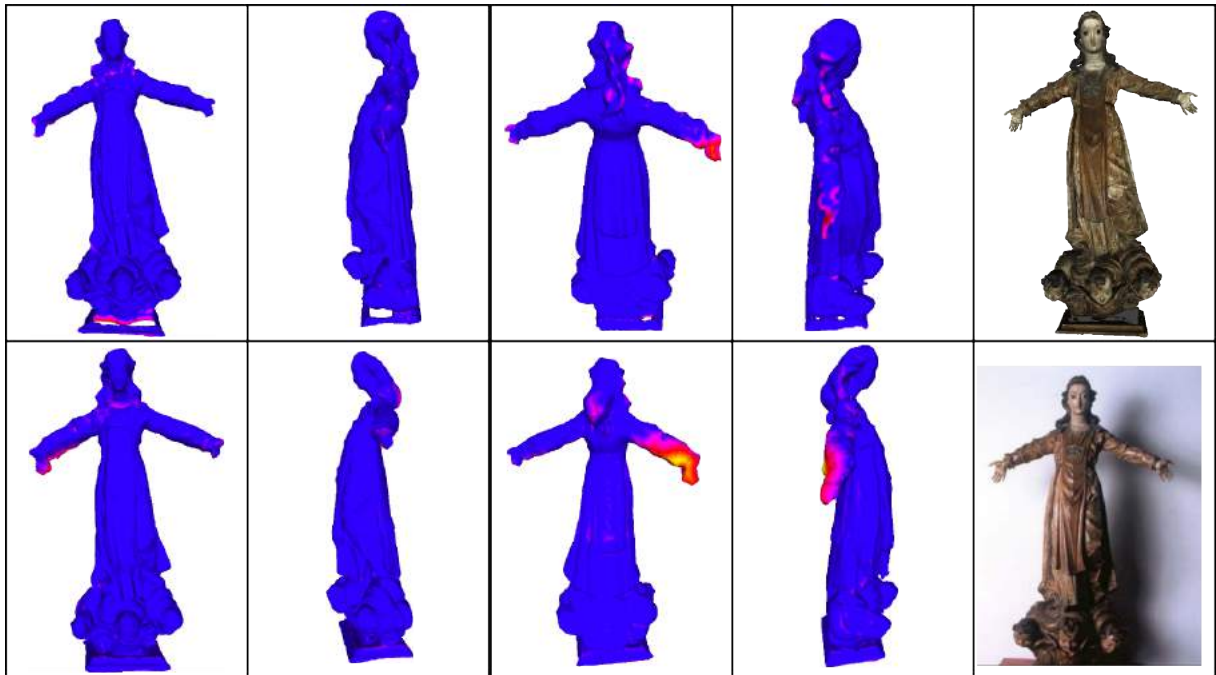


Figura 5.5: A imagem apresenta o erro relativo ao EQM ao objeto Nossa senhora das Mercês. Cui *et al.* passou pelo problema de regiões duplicadas na parte traseira do braço esquerdo da escultura.

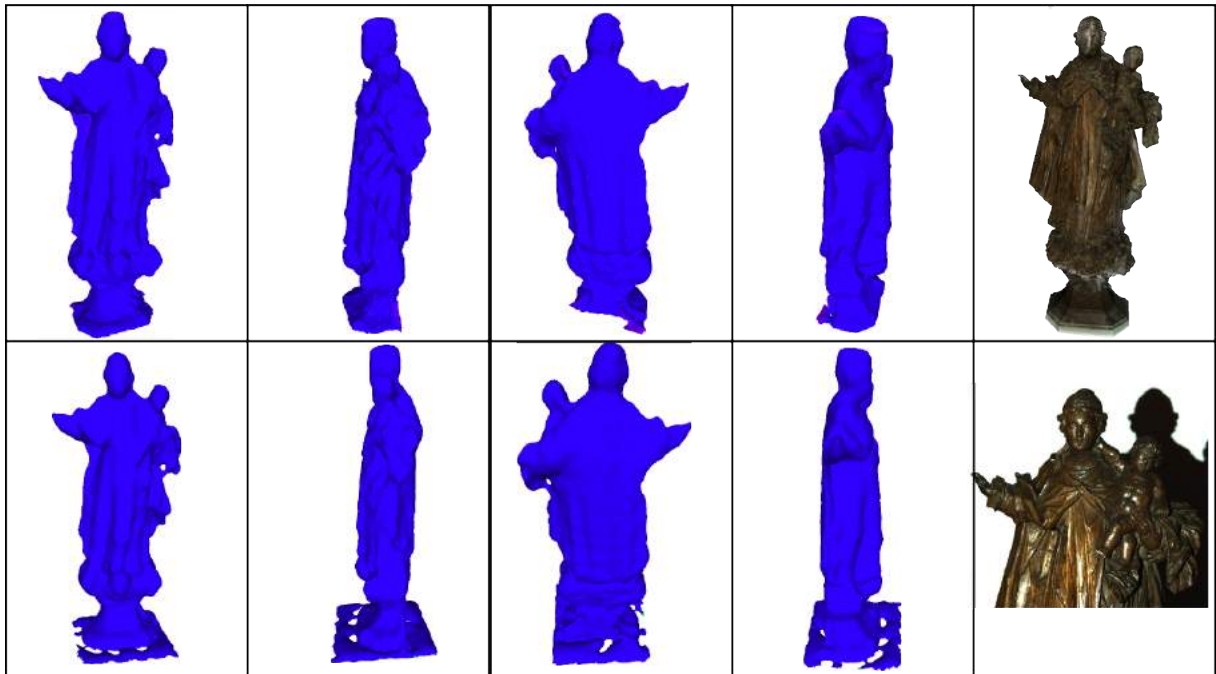


Figura 5.6: A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo 1. Visualmente os dois modelos são semelhantes.

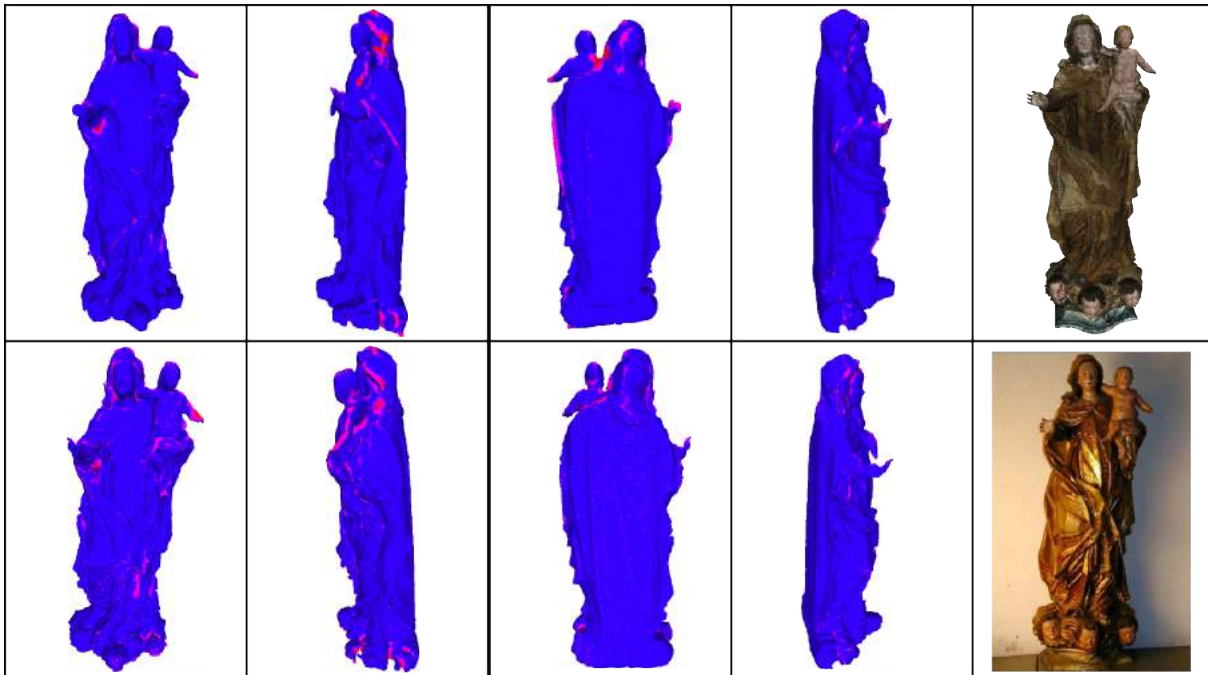


Figura 5.7: A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo 2. Visualmente os dois modelos são semelhantes.

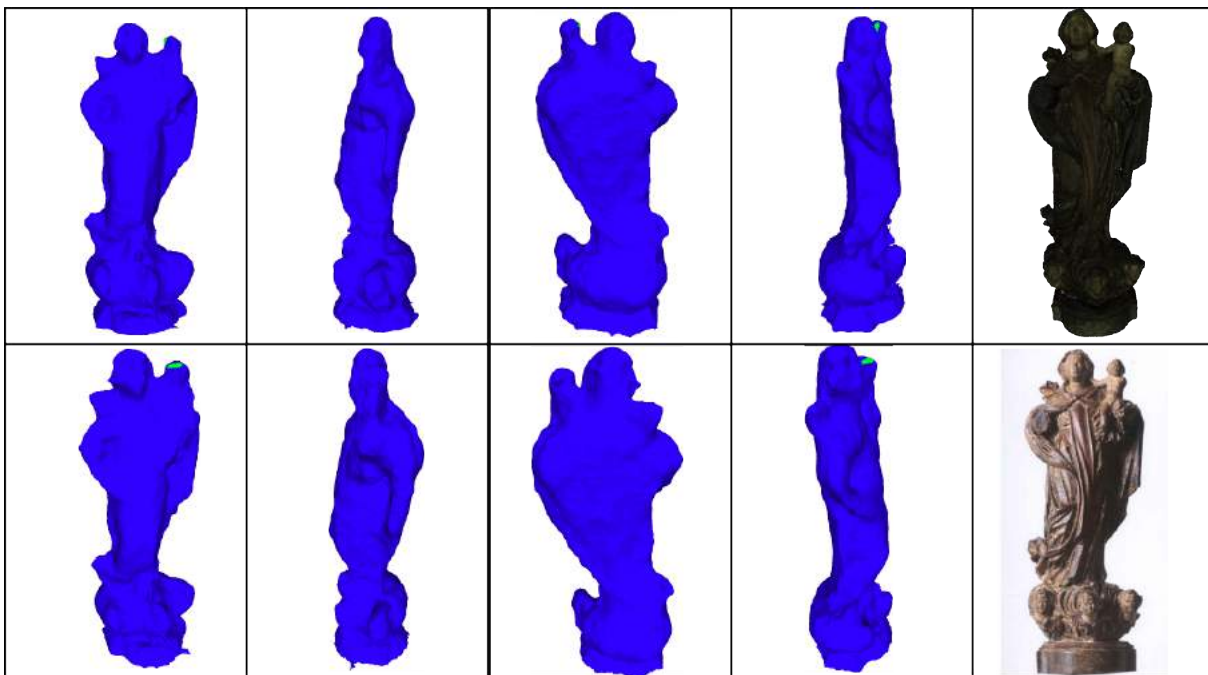


Figura 5.8: A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo 3. Visualmente os dois modelos são semelhantes.

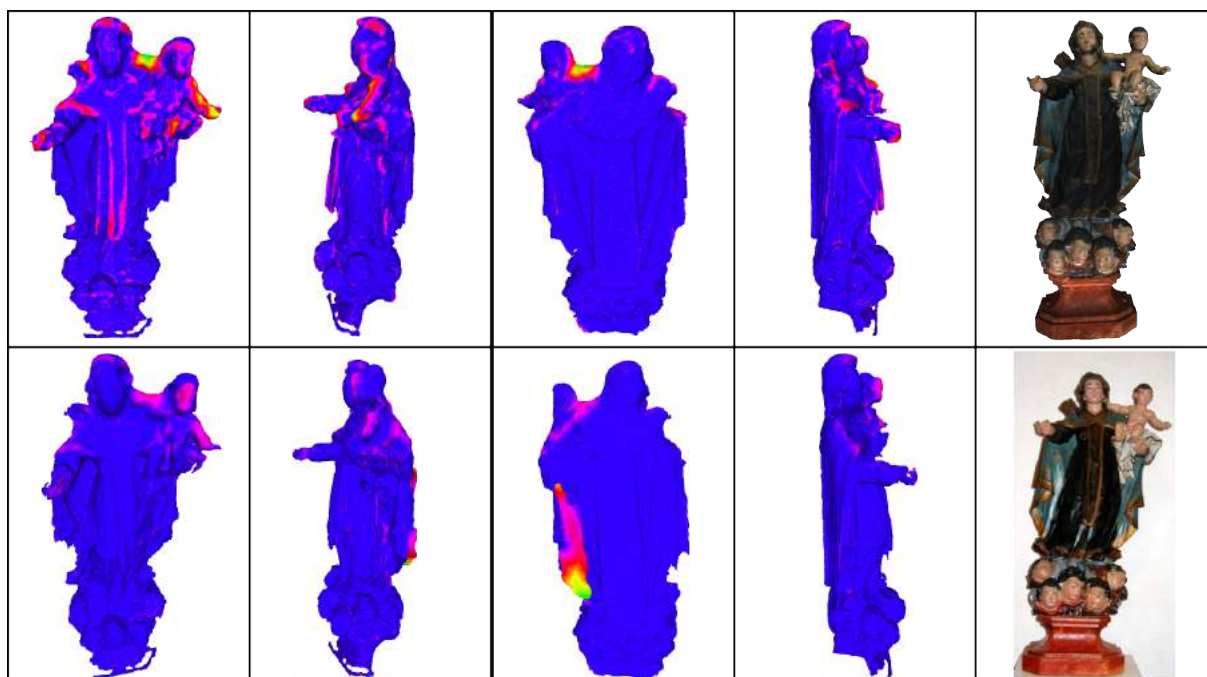


Figura 5.9: A imagem apresenta o erro relativo ao EQM ao objeto Nossa Senhora do Carmo 4. Na parte esquerda traseira do modelo gerado por Cui *et al.* existiu o problema da região duplicada.

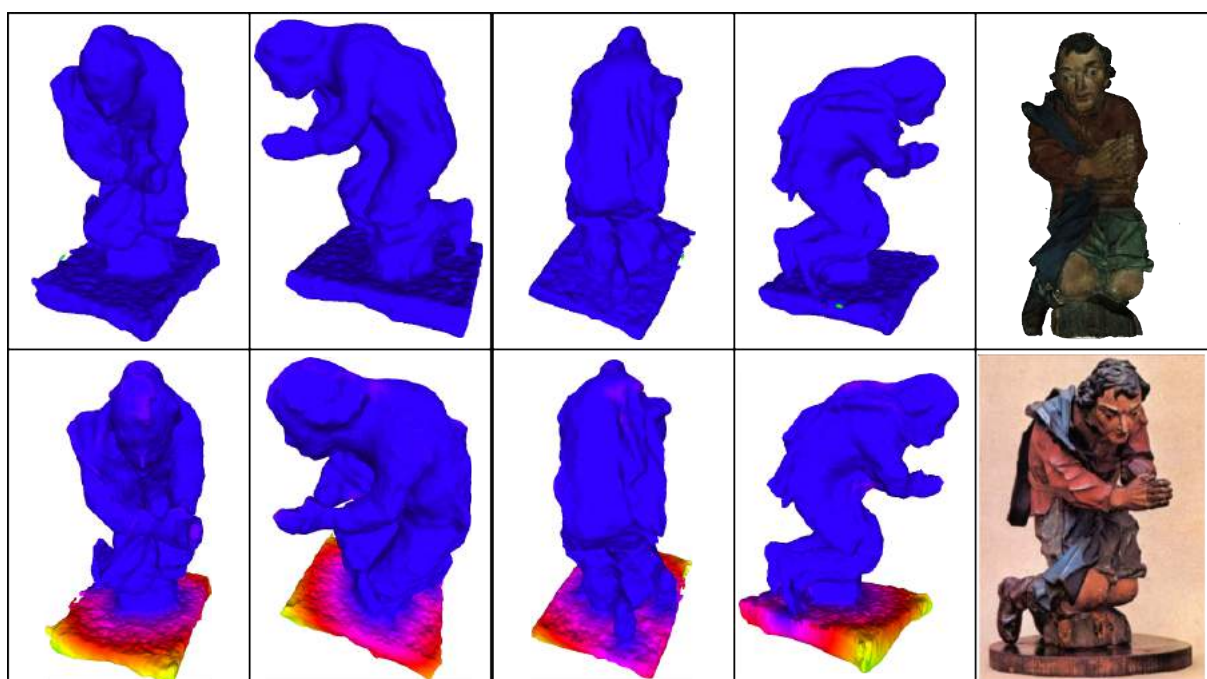


Figura 5.10: A imagem apresenta o erro relativo ao EQM ao objeto Pastor Ajoelhado. Os dois modelos são visualmente semelhantes, a diferença na base do modelo gerado por Cui *et al.* aconteceu pois este modelo é verticalmente mais alongado.

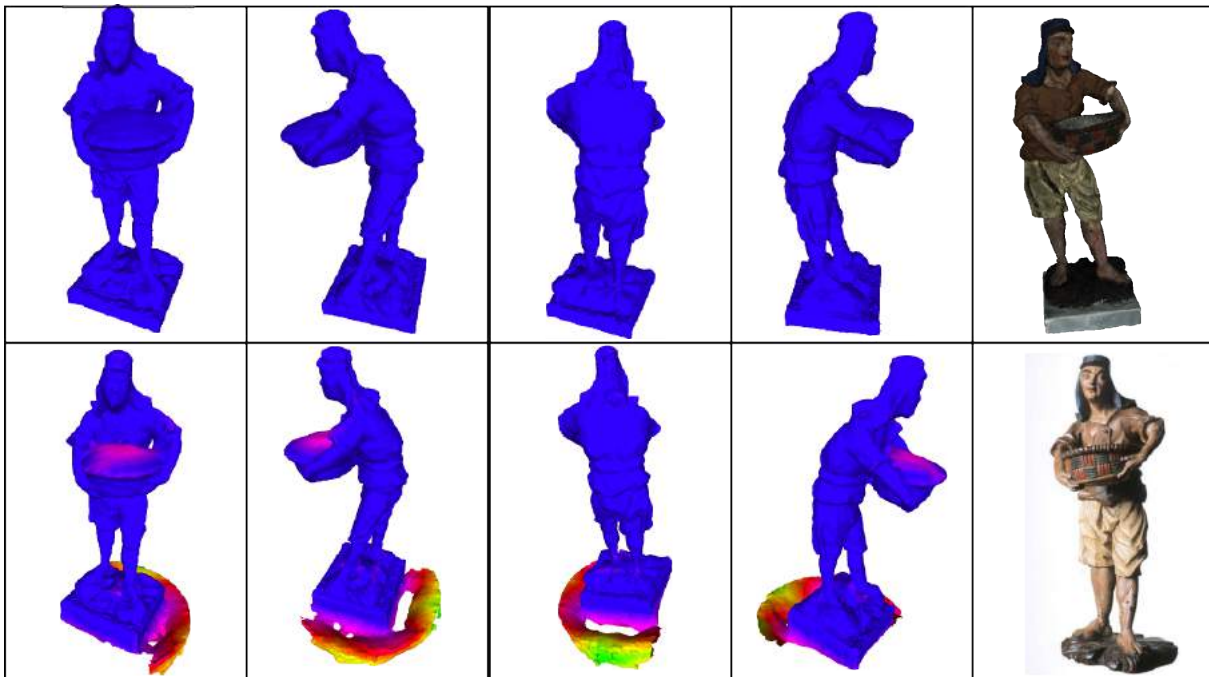


Figura 5.11: A imagem apresenta o erro relativo ao EQM ao objeto Pescador. A diferença na base do modelo de Cui *et al.* aconteceu pela região não ser adequadamente segmentada pelo método.

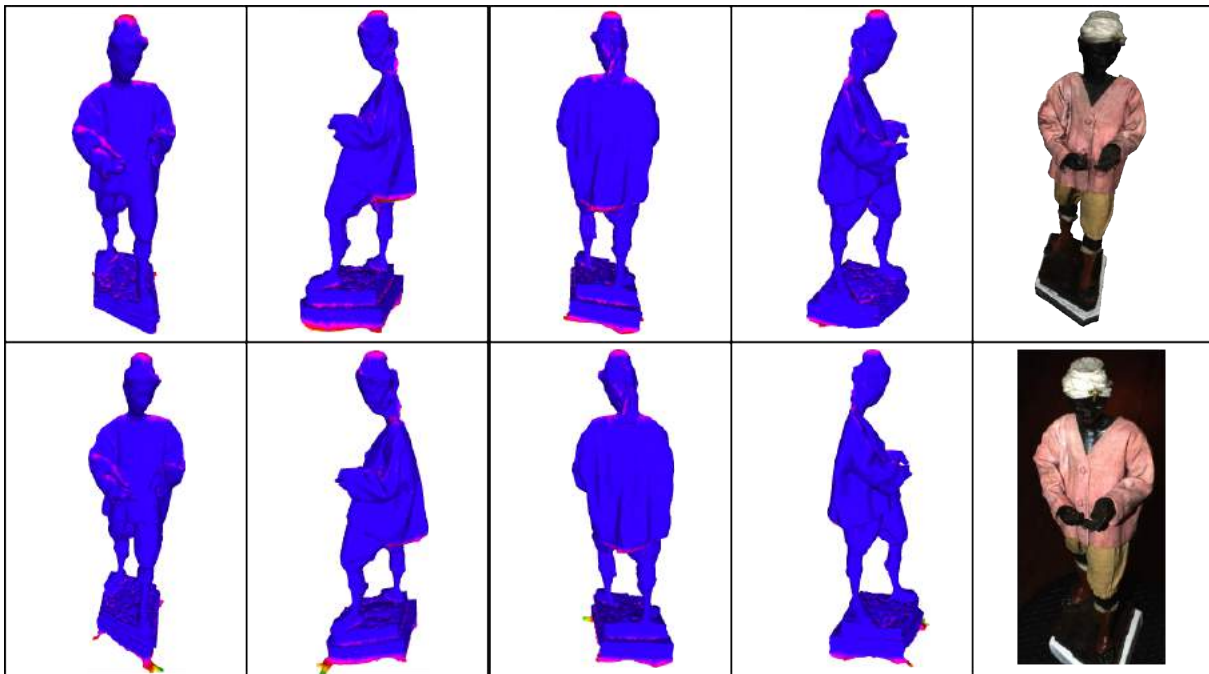


Figura 5.12: A imagem apresenta o erro relativo ao EQM ao objeto Presépio Negro. Os dois modelos são visualmente semelhantes.

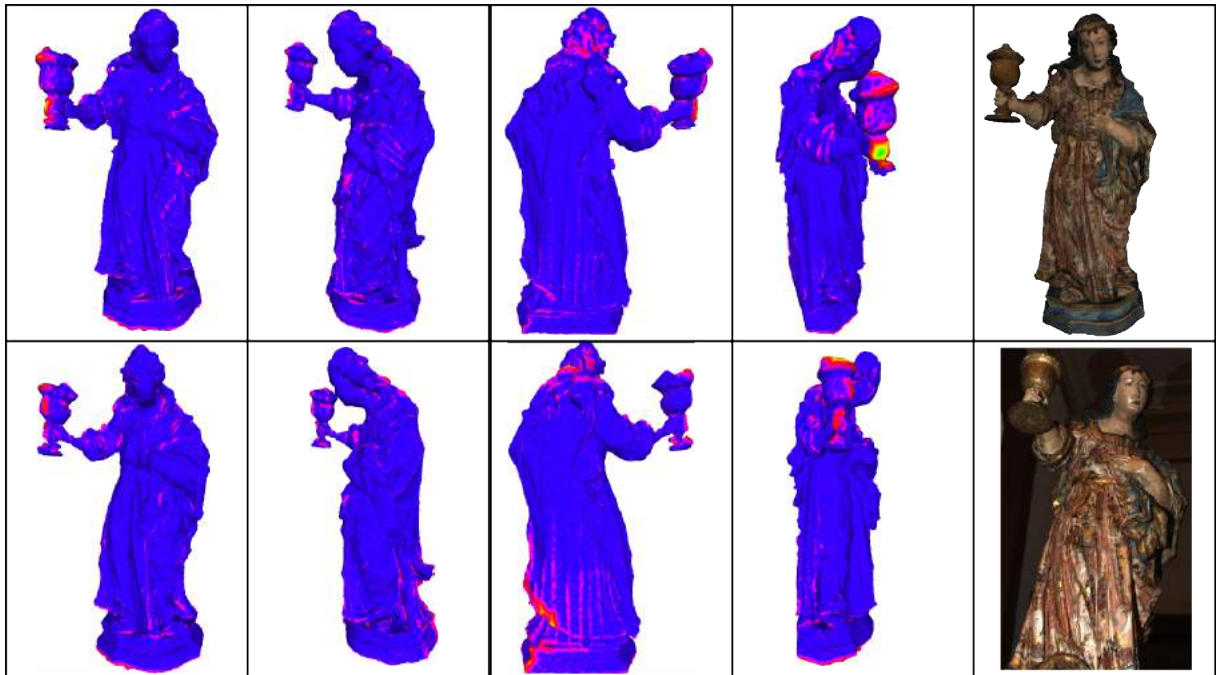


Figura 5.13: A imagem apresenta o erro relativo ao EQM ao objeto Santa Bárbara. Os dois modelos são visualmente semelhantes.

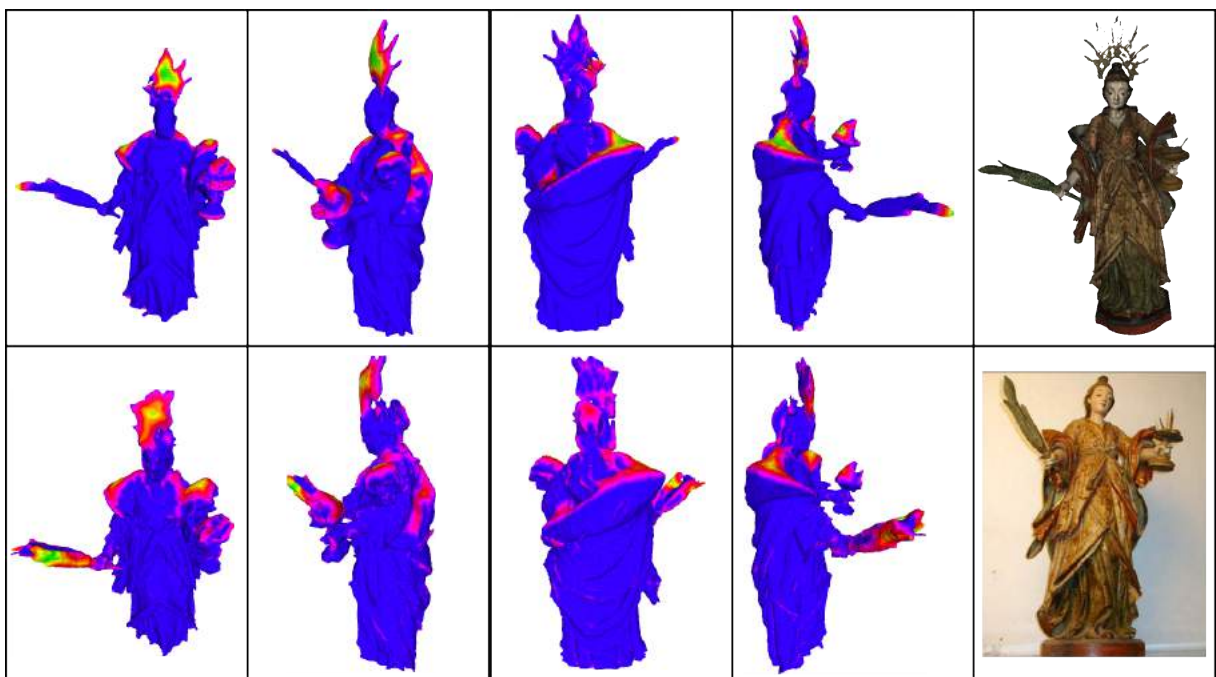


Figura 5.14: A imagem apresenta o erro relativo ao EQM ao objeto Santa Luzia. A região da coroa da escultura teve erro acentuado para os dois modelo por conta da região possuir refletância e gerar dados incorretos. O modelo gerado por Cui *et al.* também duplicou o braço direito da escultura.

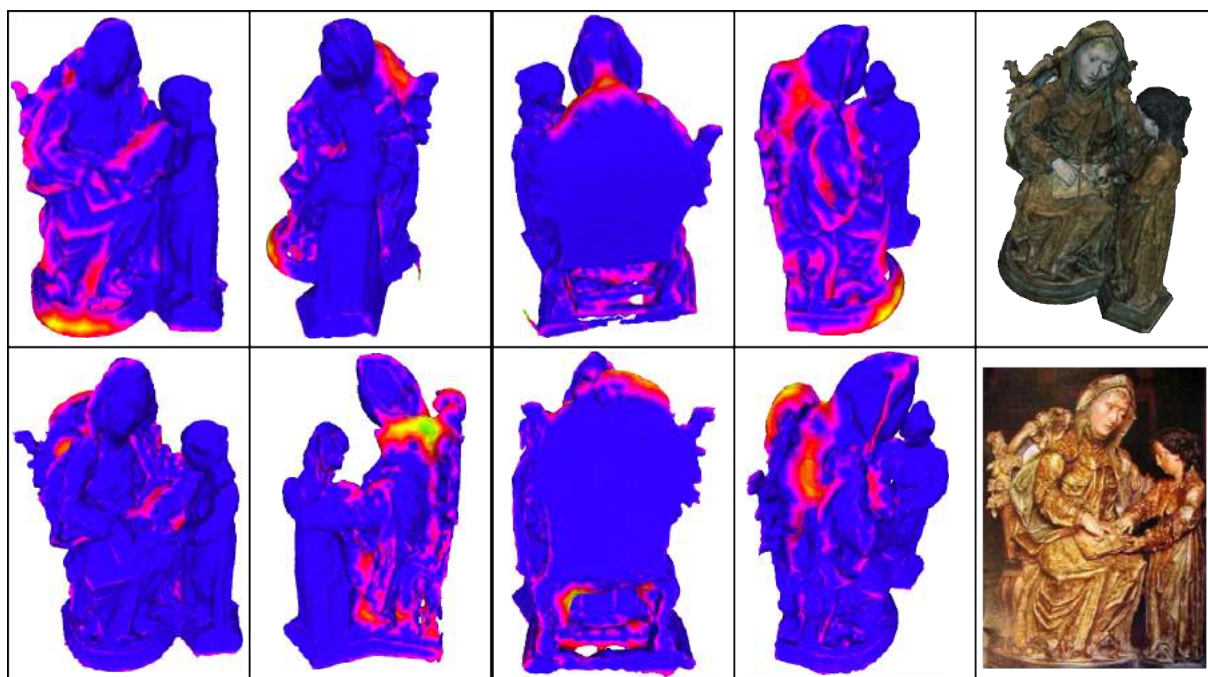


Figura 5.15: A imagem apresenta o erro relativo ao EQM ao objeto Santana Mestra 1. Os dois modelos são visualmente semelhantes.

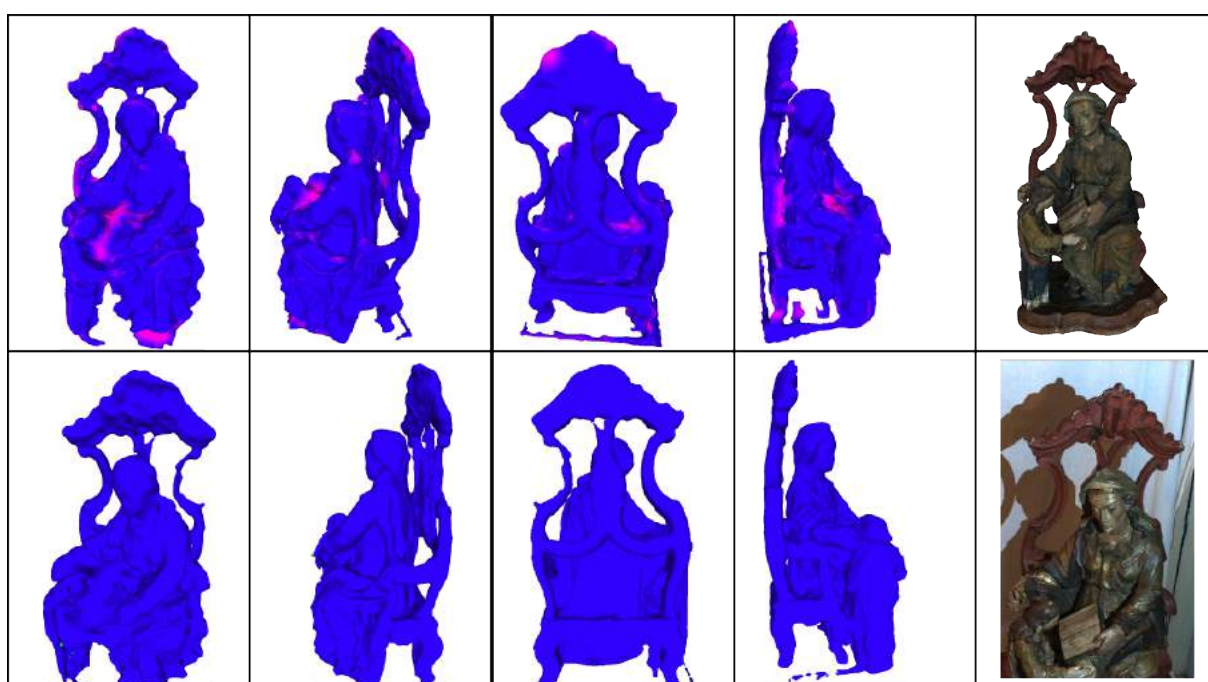


Figura 5.16: A imagem apresenta o erro relativo ao EQM ao objeto Santana Mestra 2. Os dois modelos são visualmente semelhantes.

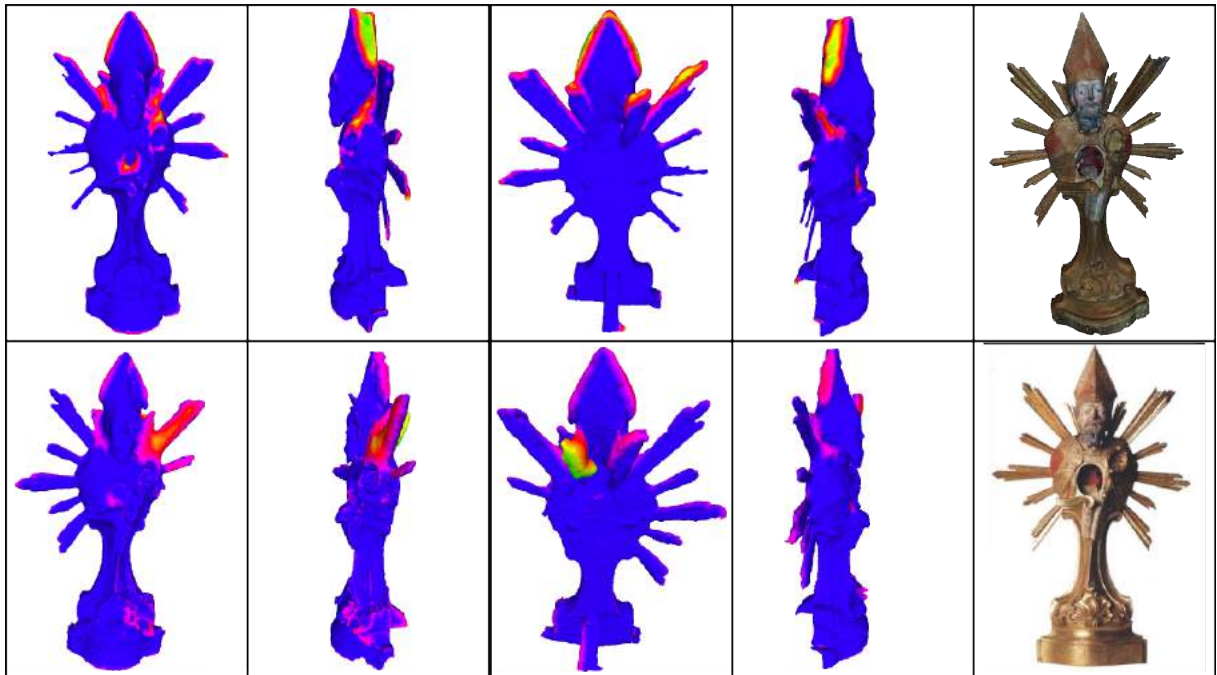


Figura 5.17: A imagem apresenta o erro relativo ao EQM ao objeto Santo Atanásio. O modelo gerado por Cui *et al.* duplicou uma pequena região na parte traseira da escultura e falhou em reconstruir o arte direita dela.

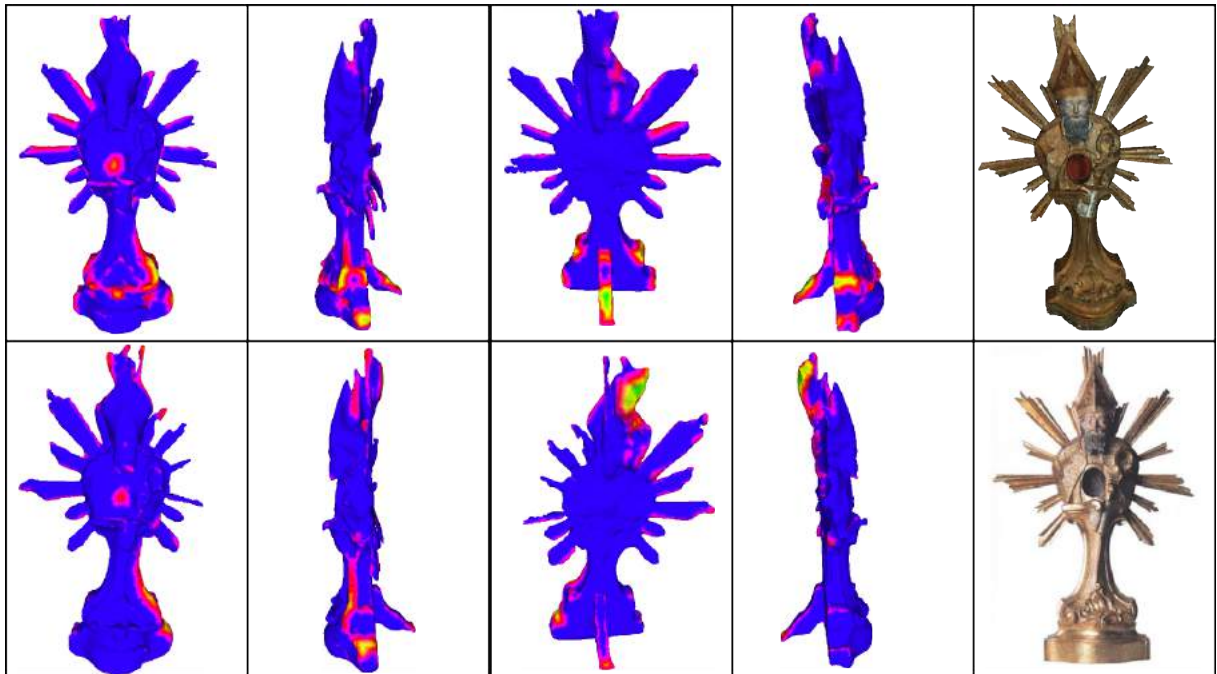


Figura 5.18: A imagem apresenta o erro relativo ao EQM ao objeto Santo Hilário. O modelo gerado por Cui *et al.* duplicou uma pequena região na parte traseira da escultura e falhou em reconstruir o arte direita dela.

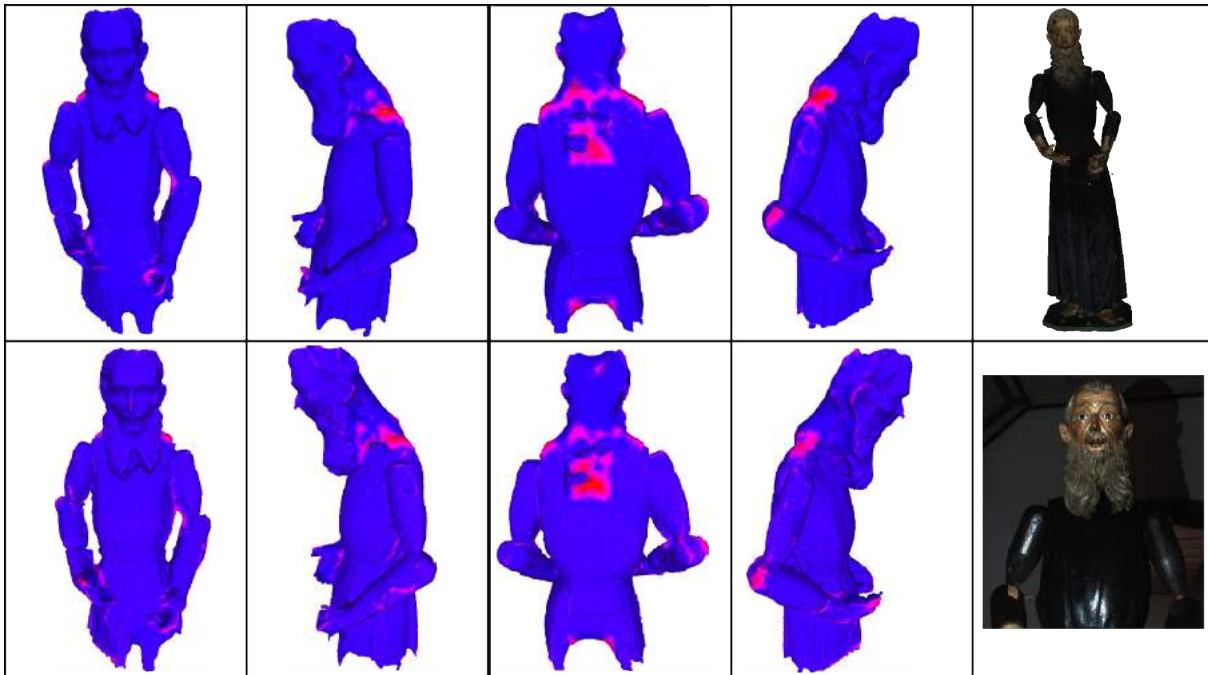


Figura 5.19: A imagem apresenta o erro relativo ao EQM ao objeto São Francisco de Paula 1. Os dois modelos são visualmente semelhantes.

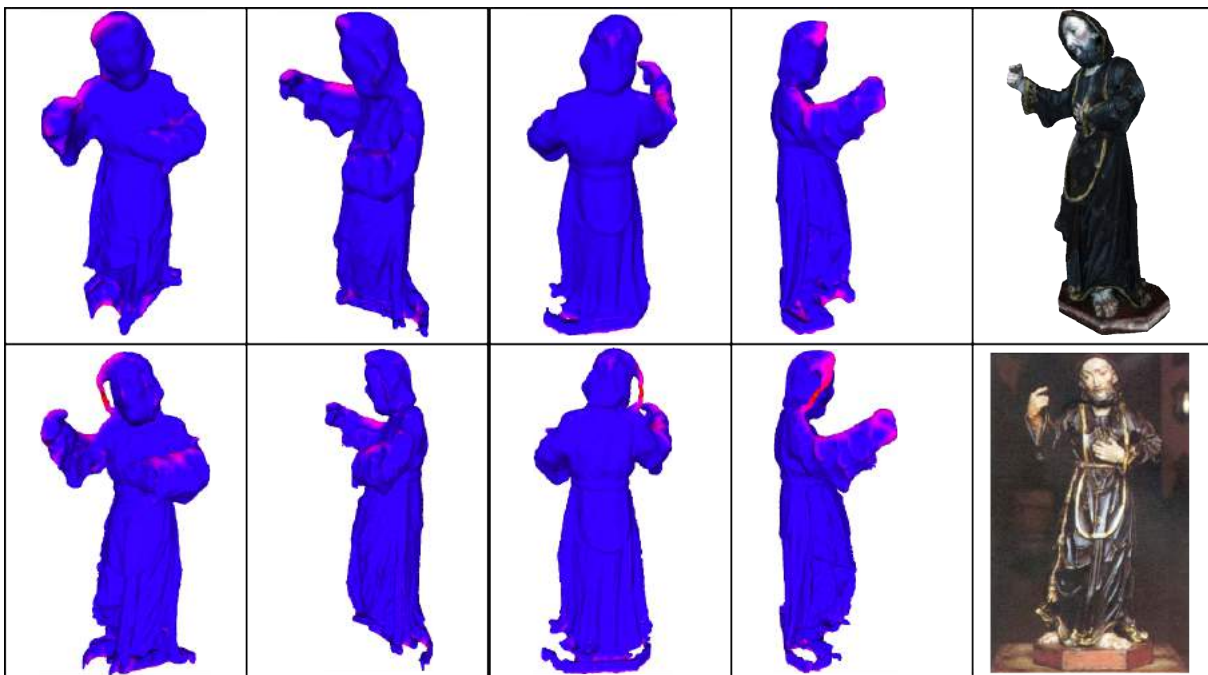


Figura 5.20: A imagem apresenta o erro relativo ao EQM ao objeto São Francisco de Paula 2. Uma pequena região faltante na parte da cabeça da escultura do modelo gerado pelo método de Cui *et al.*

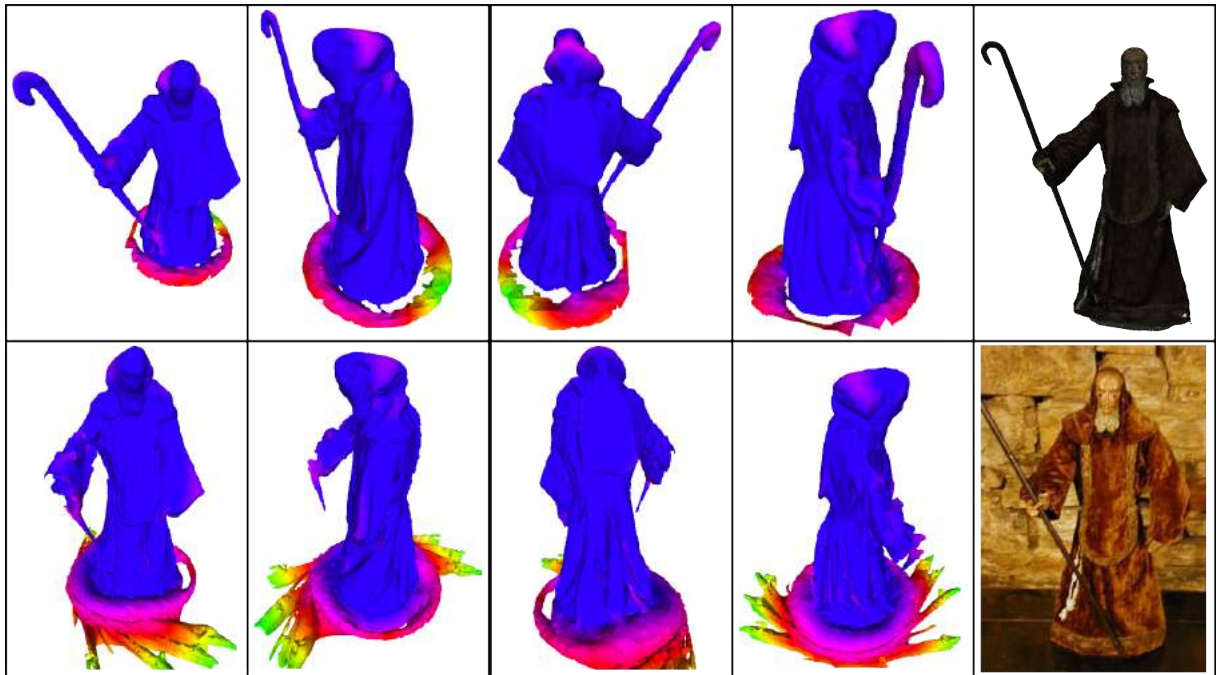


Figura 5.21: A imagem apresenta o erro relativo ao EQM ao objeto São Francisco de Paula 3. Região faltante no cajado da escultura pelo modelo gerado por Cui *et al.*



Figura 5.22: A imagem apresenta o erro relativo ao EQM ao objeto São João da Cruz. Erros ocorreram por conta das regiões reflexivas dos objetos.

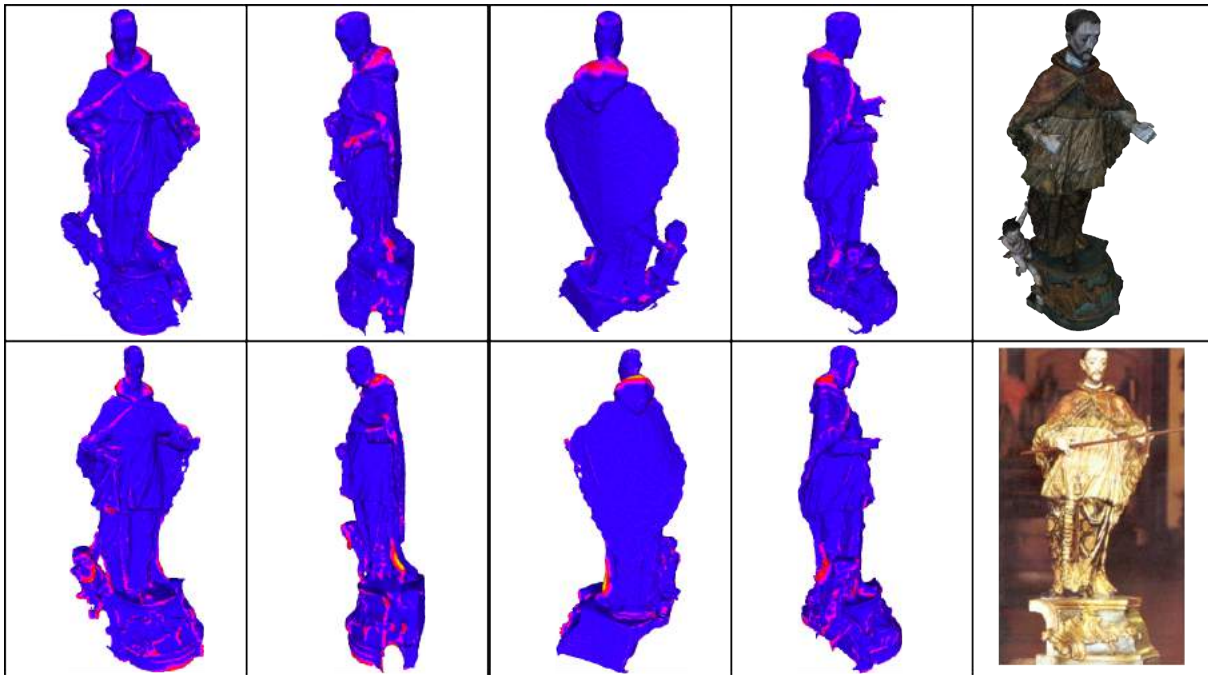


Figura 5.23: A imagem apresenta o erro relativo ao EQM ao objeto São João Nepomuceno. A parte esquerda da face da escultura é uma região faltante no modelo gerado por Cui *et al.*

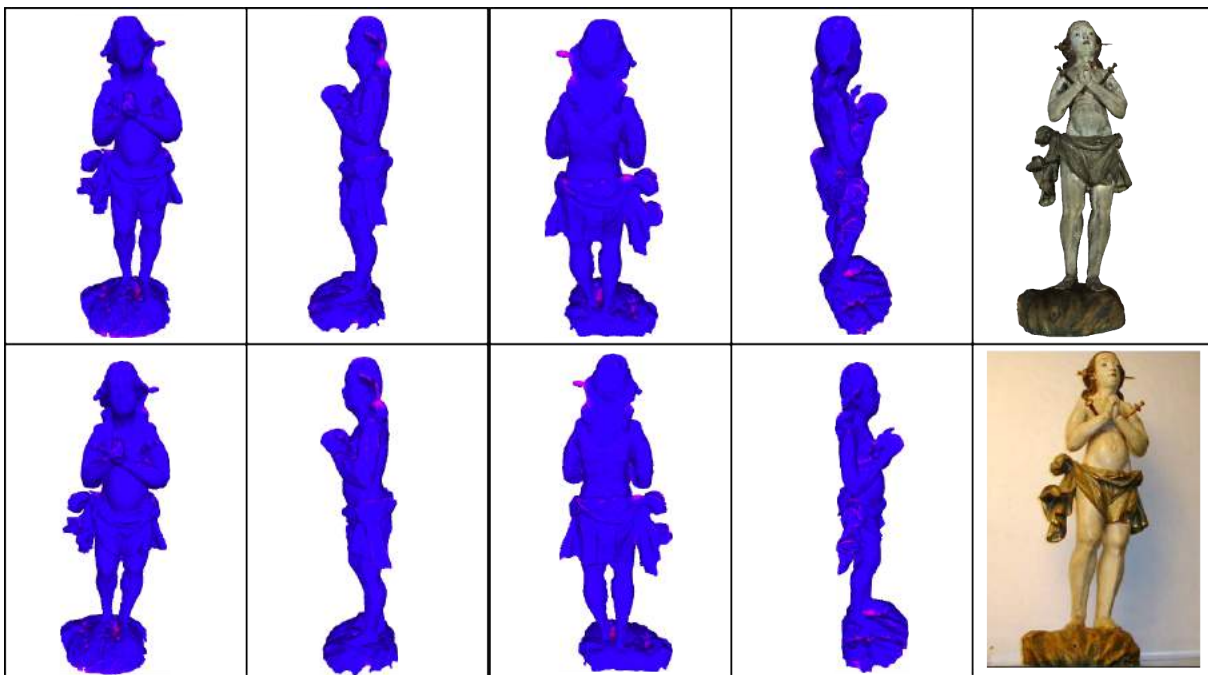


Figura 5.24: A imagem apresenta o erro relativo ao EQM ao objeto São Manoel. Os dois modelos são visualmente semelhantes.

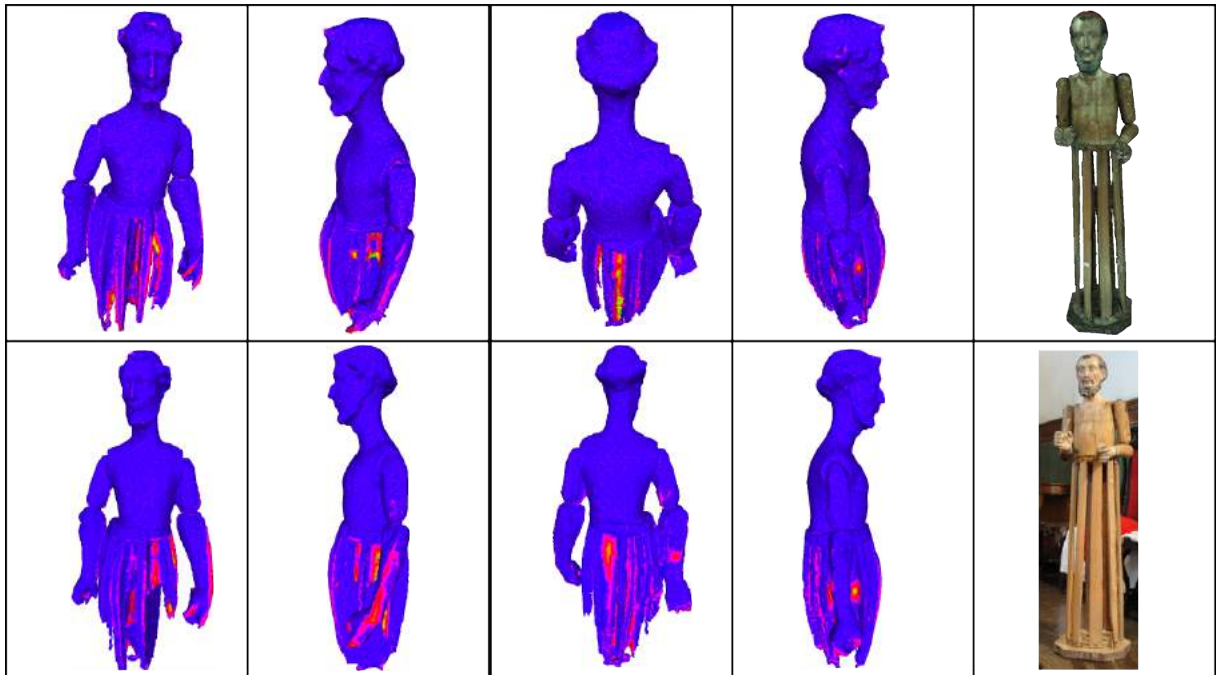


Figura 5.25: A imagem apresenta o erro relativo ao EQM ao objeto São Pedro Nolasco 1. Os dois modelos são visualmente semelhantes.

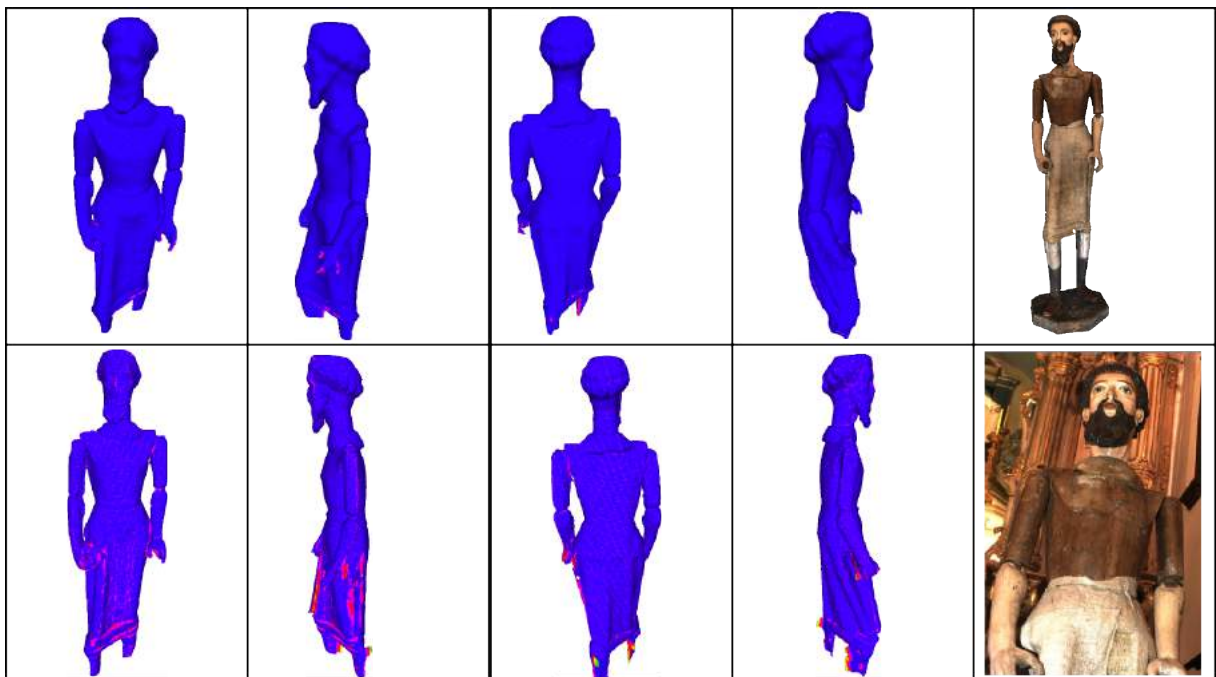


Figura 5.26: A imagem apresenta o erro relativo ao EQM ao objeto São Pedro Nolasco 2. Os dois modelos são visualmente semelhantes.

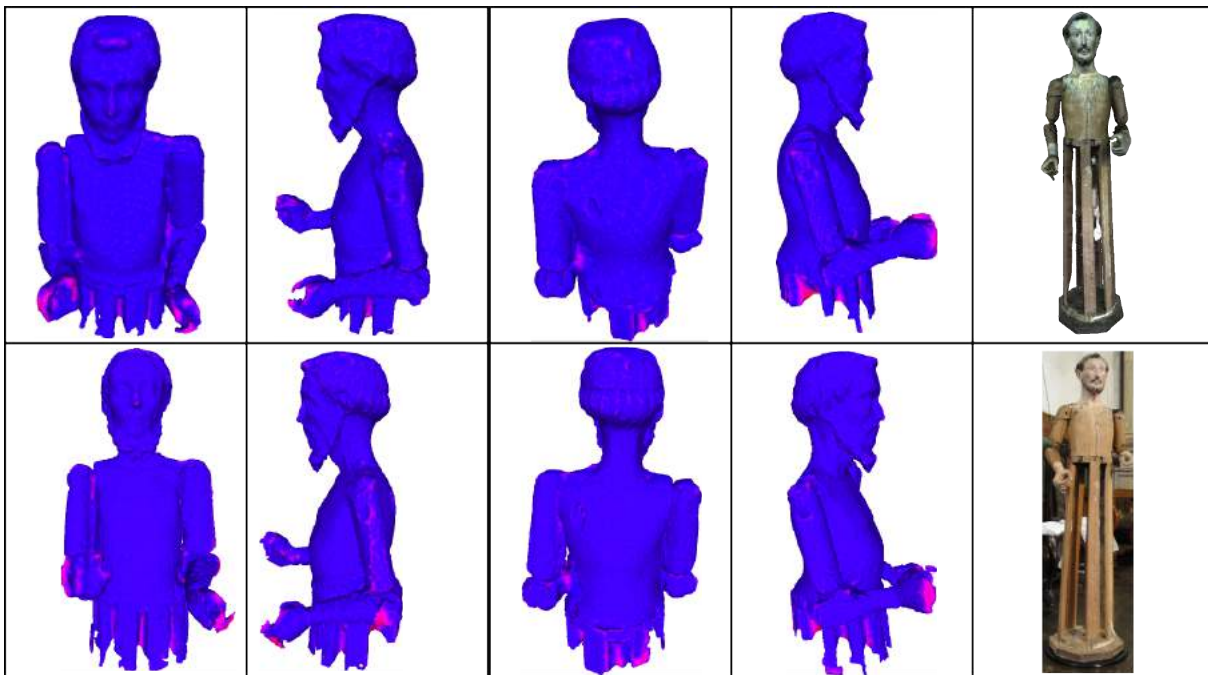


Figura 5.27: A imagem apresenta o erro relativo ao EQM ao objeto São Raimundo Nonato 1. Pequenas regiões duplicadas na mão esquerda do modelo gerado por Cui *et al.*

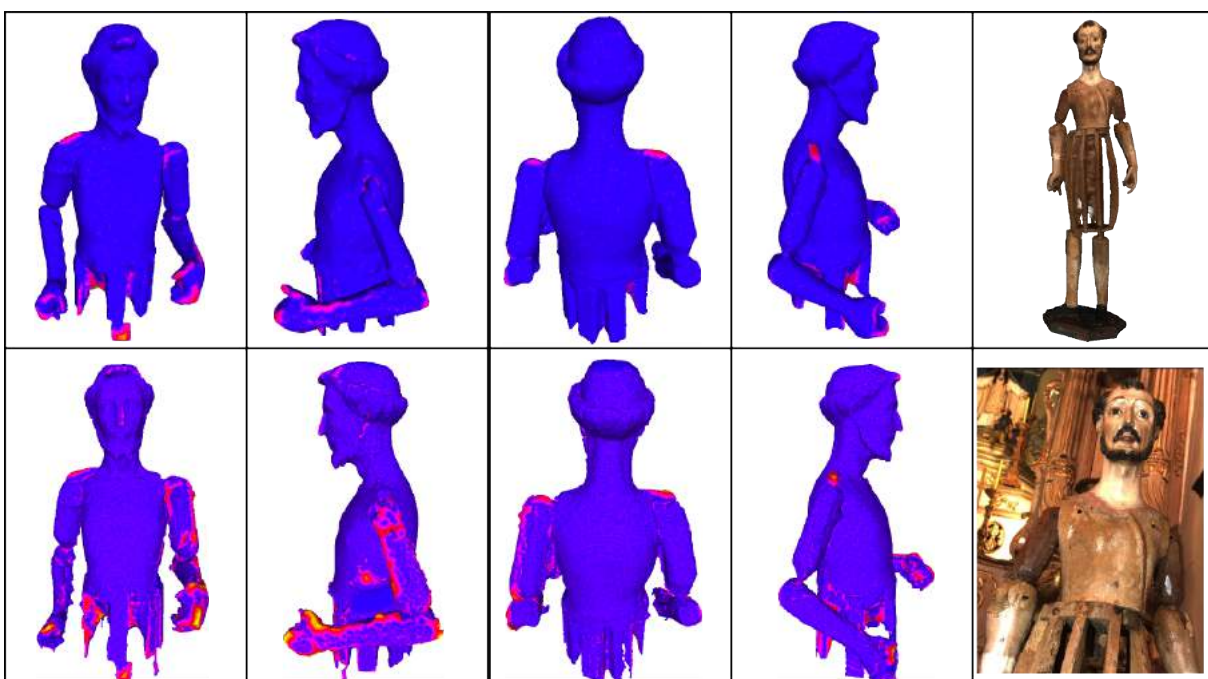


Figura 5.28: A imagem apresenta o erro relativo ao EQM ao objeto São Raimundo Nonato 2. Os modelos são visualmente semelhantes.

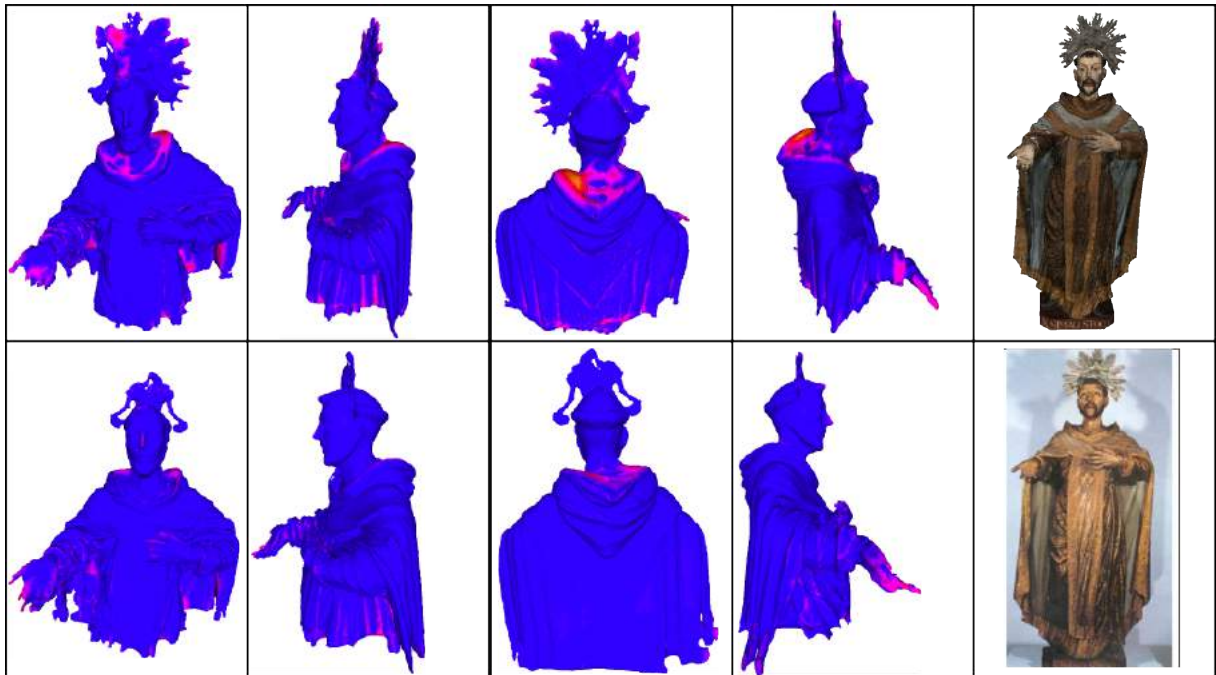


Figura 5.29: A imagem apresenta o erro relativo ao EQM ao objeto São Simão Stock. Problemas causados pela reflexividade do objeto no topo da cabeça da escultura.

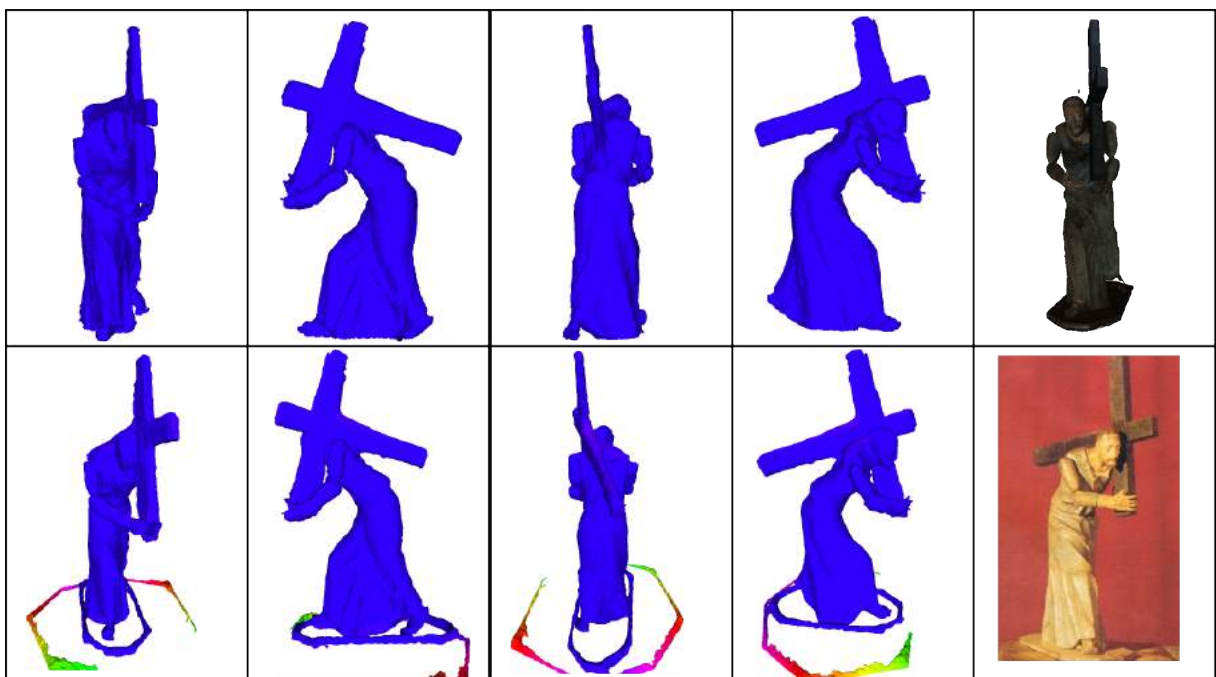


Figura 5.30: A imagem apresenta o erro relativo ao EQM ao objeto Senhor dos Passos. No modelo gerado por Cui *et al.*, algumas regiões na base não foram segmentadas pelo método.

5.3 Discussão sobre o experimentos

Nos experimentos o método proposto apresentou resultados favoráveis em relação com os dois outros métodos avaliados. O diferencial foi a etapa de alinhamento e a estratégia de tratamento de *loop closure* que permitiu modelos sem as deformações causadas por regiões duplicadas ou faltantes. No comparativo quantitativo feito pela distância de Hausdorff foi adotado o modelo de precisão submilimétrica como o padrão ouro. O método proposto apresentou erro menor para 27 das 30 obras avaliadas. Nos 3 objetos em que Cui *et al.* foi melhor sucedido (Figuras 5.3, 5.8 e 5.25 obtiveram geometrias muito semelhantes ao método proposto, sem distinção visual significativa mas com erro médio menor porém semelhante.

O método de Cui *et al.* e o método proposto possuem tempos de execução muito próximos, com pequena vantagem para o método proposto, as implementações não focaram em otimizações então tempos ainda melhores podem ser obtidos. No trabalho proposto em média a etapa de alinhamento consumiu 40% do tempo enquanto a integração volumétrica os outros 60%. Além da aquisição de dados através do SmartScan ser trabalhosa e lenta, as malhas capturadas necessitam ser pré-alinhadas através de marcações feitas manualmente. As demais etapas possuem tempo de processamento semelhante ao método proposto utilizando Kinect.

Os modelos gerados com smartScan possuem definição geométrica superior. Porém certas aplicações não requerem precisão submilimétrica. Por exemplo a visualização em museus virtuais, que podem combinar modelos gerados pelo Kinect com a textura de câmeras externas para resultados visuais mais verossímeis.

5.4 Comparativo

Como já foi citado, um dos objetivos deste trabalho é apresentar uma discussão e comparação entre a qualidade dos sistemas que utilizam scanners de precisão submilimétrica e avaliar a viabilidade de utilizar câmeras RGB-D, de maior velocidade, em obras de arte. A digitalização do acervo cultural foi feita utilizando o SmartScan da Breuckmann e o Microsoft Kinect. Na sequência é apresentado uma discussão para essa experiência com cada aparelho.

5.4.1 SmartScan

Nos experimentos apresentados foi utilizado o Smart SCAN 3D-HE da Breuckmann¹ cuja tecnologia de luz estruturada é semelhante aquela aplicada pelo Kinect. O Breuckmann é composto de um sistema calibrado composto por duas câmeras e um projetor que emite um conjunto de padrões de luz sobre a superfície do objeto, pela deformação do padrão de luz combinado com o deslocamento de pontos correspondentes entre as duas câmeras uma imagem de profundidade é calculada. A precisão depende do conjunto de lentes utilizada e pode chegar na casa dos micrômetros, segundo as especificações do fabricante. As mesmas técnicas de alinhamento e integração que foram aplicadas aos dados do Kinect também foram adotadas nos modelos gerados neste trabalho utilizando o SmartScan, mas sem o pré-processamento e com uma etapa extra de pré-alinhamento semi-assistido.

A precisão dos dados do smartScan dispensa qualquer pré-processamento nos dados e o efeito de *loop closure* causado pelo acúmulo de erro de alinhamento é reduzido proporcionalmente ao aumento da qualidade e diminuição do número de malhas. No entanto a grande dificuldade está na etapa de aquisição. Uma única captura pode necessitar de até 1 minuto e meio para ser gerada dependendo da configuração adotada, o scanner precisa estar totalmente estático durante esse período, por isso também é necessário utilizar um tripé para a aquisição. Entre cada captura é necessário realizar o reposicionamento do equipamento e objeto, que pode demandar diversos minutos dependendo das condições e ponto de vista desejado. As campanhas de digitalização de acervos necessitam do deslocamento de uma equipe para o local com até centenas de kilos em equipamentos entre, iluminação, fontes de energia, computadores, tripés, bases giratórias. Além de bloquear o acesso a visitação das peças pelo período da digitalização. Por isso o tempo de aquisição nestas missões é muito importante. Na Figura 5.31 é apresentada uma imagem que ilustra essa situação.

¹<http://aicon3d.com/products/breuckmann-scanner/smartscan>



Figura 5.31: A foto apresenta o momento da aquisição pelo sensor SmartScan 3D - Breuckman. Ele está na altura de 3 metros com auxílio de um tripé. A escultura em madeira representa São Pedro Nolasco e está localizada na igreja de Nossa Senhora do Carmo em Ouro Preto - MG.

5.4.2 Kinect

Os problemas logísticos durante a aquisição são minimizados ao utilizar câmeras RGB-D. Primeiramente a aquisição é de 10 a 20 vezes mais rápida, favorecendo acervos maiores. Qualquer usuário sem treinamento também é capaz de operar o sistema pois assemelha-se a uma câmera filmadora comum e a resposta em tempo-real permite identificar a conclusão da aquisição com facilidade. Os equipamentos são leves, portáteis e necessitam apenas de um notebook para serem operados. O custo é centenas de vezes menor do que os scanners com a precisão do SmartScan. Portanto uma opção muito mais viável financeiramente e logisticamente.

A decisão sobre que tipo de equipamento adotar depende de diversas questões: Quantidade de recurso existente para o projeto; disponibilidade de tempo do acervo e operadores; tamanho do acervo; topologia e materiais que compõe o acervo; precisão necessária para o objetivo final dos modelos. Portanto cada projeto terá um tipo de sensor mais indicado.

5.4.3 Intel Realsense F200

Além do Kinect foram realizados experimentos utilizando o sensor Intel Realsense F200. Este sensor utiliza a mesma tecnologia do Kinect, luz estruturada, porém ele possui um conjunto de lentes e especificações gerais diferentes. O objetivo primário deste sensor é reconhecimento facial e captura de gesticulação manual. Já o Kinect é originalmente pensado para reconhecimento de gesticulação corporal. O sensor F200 possui menor abertura angular em suas lentes e consegue operar na distância de 0.2 até 1.2 m, contra os 1m até 10m do Kinect. No comparativo com o Kinect esta diferença faz com que o F200 seja mais preciso na digitalização de objetos menores porém ineficaz quando se deseja operar em uma distância maior do que 1,2 metros. Na Figura 5.32 uma escultura com 15cm de altura foi digitalizada com diferentes sensores e ilustra essa vantagem. O sensor F200 só foi adquirido depois de terminada a missão de aquisição do projeto Aleijadinho Digital, por isso não há dados daquelas obras com este sensor. A tabela 5.2 apresenta o comparativo em EQM entre o modelo gerado a partir do objeto original com um segundo modelo criado a partir de uma impressão 3D deste primeiro modelo. Este experimento foi realizado visando avaliar a repetibilidade e precisão dos três sensores.

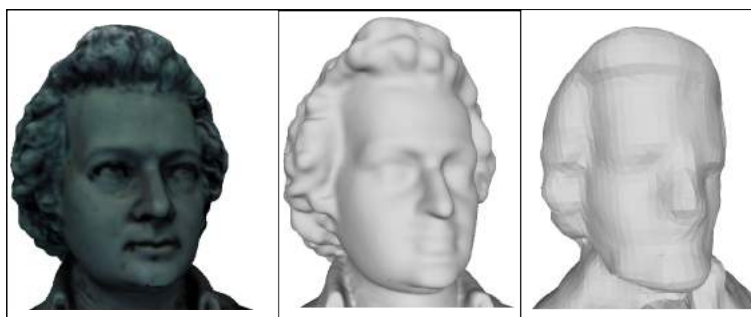


Figura 5.32: A imagem apresenta modelos 3D de uma escultura do compositor Mozart. A primeira imagem é o modelo gerado com o Intel Realsense F200 com textura utilizando o método proposto. A segunda imagem é a renderização do mesmo modelo sem textura. A terceira imagem é o mesmo objeto gerado utilizando o método proposto porém capturado utilizando o Microsoft Kinect.

Experimento de rescan	Resultado comparativo RMS
SmartScan	0.186731 mm
Realsense f200	1.099213 mm
Kinect	2.348048 mm

Tabela 5.2: Tabela comparativa apresentado o EQM entre os modelos de buda comparados com modelos de sua impressão 3D.

5.4.4 Scanner de tempo de percurso

Também foram realizados experimentos utilizando os dados adquiridos com um scanner por tempo de percurso MESA Swissranger SR4000 da base de dados apresentada em Cui *et al.* [Cui et al., 2010]. A base é composta por 5 objetos de aproximadamente 20 cm cada um. Com 600 imagens por objeto. A aquisição não foi feita 360 graus, mas apenas malhas 3D frontais dos objetos, como padrão ouro é apresentado o modelo digitalizado utilizando um scanner laser Vivid Minolta 910. A Figura 5.33 apresenta os objetos que compõem a base. Na Tabela 5.3 foi comparado o EQM do método proposto com Cui *et al.*. Os resultados foram muito semelhantes pois não existiu o efeito de acúmulo de erro de alinhamento nesta base visto que as imagens foram frontais.



Figura 5.33: A figura mostra fotos dos objetos que compõe a base de dados disponibilizada por Cui *et al.* [Cui et al., 2010].

Modelos	Cui et al.	Método proposto
Head	2.36144	2.011378
Vase	4.40203	4.185429
Angels_handheld	3.845092	3.440284
Buddha	4.38733	3.918021
Angels	4.40203	4.345893
Média	3.8795844	3.580201

Tabela 5.3: Tabela comparativa o EQM entre os modelos gerados com o método proposto e Cui *et al.* em relação ao modelo gerado com um scanner de precisão submilimétrica.

Capítulo 6

Conclusão

Esta tese apresentou uma pesquisa relacionada a reconstrução de modelos 3D utilizando câmeras RGB-D, com foco na preservação digital de acervos culturais. No começo foi apresentado os principais projetos neste campo da preservação digital, seus desafios e soluções. Na sequência foram avaliadas as tecnologias e sensores para geração 3D e suas vantagens individuais. Foi apresentado uma discussão entre os benefícios da velocidade, praticidade, logística entre outros ao adotar câmeras RGB-D, e.g. Kinect contra a qualidade geométrica de um scanner de maior precisão como o SmartScan da Breuckmann. Em seguida foi apresentado os principais trabalhos estado-da-arte que utilizam câmeras RGB-D em relação ao trabalho proposto. Os trabalhos foram divididos em tempo real e *offline* e foi discutido o comparativos entre as duas estratégias.

Finalmente foi apresentado o método proposto em si. A abordagem foi dividida em quatro etapas principais:

- Captura dos dados, onde foi discutida a forma de captura adotada e as propriedades específicas dos sensores utilizados nos experimentos, Kinect e Intel Realsense.
- Pré-processamento, onde foi aplicado o filtro bilateral de Matsuo *et al.* [Matsuo et al., 2013] para atenuar ruídos e inferir os dados tirados de regiões sem informação pelo sensor incrementando de forma geral os dados de entrada.
- Alinhamento, onde todas as capturas 3D são transformadas para um sistema de coordenadas comum. Foi adotada a estratégia de alinhamento em pares baseada em ICP. E aplicada em

conjunto com método de Pulli [Pulli, 1999] uma metodologia original que ao expandir o grafo de conectividade diminui o efeito do acúmulo de erro, *loop closure*, responsável por deformações nos experimentos avaliados com os métodos estado-da-arte.

- Integração, responsável por transformar as capturas 3D já alinhadas em um modelo 3D final único.

Por fim foi apresentado os resultados experimentais. O grupo de pesquisa IMAGO, do qual os autores deste trabalho fazem parte, está envolvido em um projeto de preservação digital do acervo deixado pelo artista barroco Aleijadinho, foram digitalizadas mais de 30 obras de arte presentes em museus e igrejas da região de Ouro Preto - MG. Estas esculturas em madeira pintada do século XVIII possuem formas e tamanhos variados e foram adotadas como teste experimental para o método aqui proposto. O método foi comparado com outras duas abordagens de maior destaque na literatura pela precisão em reconstruções 3D utilizando câmeras RGB-D, Cui *et al.* [Cui et al., 2013] e *Kinect Fusion* [Izadi et al., 2011]. O método proposto apresentou resultados favoráveis quanto a qualidade geométrica dos modelos. Esse ganho na qualidade se deve principalmente ao efeito *loop closure* que foi atenuado pelo método original apresentado na etapa de alinhamento.

Para validar a hipótese apresentada na introdução foram oferecidos três **objetivos** principais:

1. Realizar uma revisão sobre os principais métodos e projetos de digitalização 3D de acervos culturais, em especial adotando câmeras RGB-D
2. Apresentar um comparativo qualitativo e quantitativo do método proposto contra o atual estado-da-arte
3. Apresentar uma discussão e comparação entre a qualidade dos sistemas que utilizam scanners de precisão submilimétrica e avaliar a viabilidade de utilizar câmeras RGB-D, de maior velocidade, em obras de arte.

O objetivo 1 é importante para entender o contexto atual da digitalização de acervos culturais e suas limitações e foi abordado no Capítulo 2. A conclusão obtida é que os sensores

tradicionalmente utilizados nesta etapa tornam o processo de aquisição caro e trabalhoso. As câmeras RGB-D são uma alternativa, embora possuam menor precisão nos dados.

O objetivo 2 foi abordado no capítulo 5. No comparativo quantitativo pela distância de Hausdorff adotando o modelo de precisão submilimétrica como *ground truth* o método proposto apresentou um resultado melhor para 27 das 30 obras avaliadas contra o métodos estado-da-arte. No comparativo qualitativo é possível identificar que deformações que duplicam ou omitem regiões não afetaram o método proposto tanto quanto afetou os métodos estado-da-arte. Isso aconteceu principalmente em função do método de alinhamento global, uma contribuição original deste trabalho.

O objetivo 3 também foi abordado no capítulo 5. A conclusão é de que a precisão dos modelos adotando Kinect está na ordem de 1 4 milímetros e os scanners tradicionais, tal qual SmartScan, possui precisão submilimétrica. Os sensores RGB-D evoluem em precisão a cada novo modelo lançado, e mesmo com a qualidade atual já é possível utilizar os modelos gerados em aplicações ligadas a digitalização de acervos culturais tal qual visualização em museus virtuais.

Os objetivos desta pesquisa foram concluídos o que leva a entender como válida a **hipótese**: através do método proposto é possível realizar reconstruções 3D utilizando câmeras RGB-D com maior precisão do que o estado-da-arte até então. Este método viabiliza a utilização de câmeras RGB-D para diversas aplicações dentro do contexto de preservação digital de acervos culturais.

6.1 Trabalhos Futuros

Entre as possíveis melhorias para o trabalho proposto é possível destacar quatro:

1. **Segmentação automática do objeto na cena.** O método proposto aplicou apenas métodos básicos de limiarização, por ser simples de implementar e aplicável em diversas situações. No entanto existem na literatura diversas técnicas de segmentação aprimoradas utilizando sensores RGB-D. O próprio KinectFusion sugere segmentação do objeto baseado na movimentação quando este se move e o fundo não. Aprimoramentos na

segmentação permitiria maior flexibilidade na captura, sem necessidade de intervenções manuais para remoção do fundo.

2. **Método para recuperar referencial quando objeto não está na cena.** É assumido durante os experimentos que durante toda a captura o objeto está sempre em cena e existe sobreposição entre imagens sequenciais. E quando esse não é o caso o método proposto não possui uma forma de recuperação automática do referencial, e se torna necessário uma intervenção manual apagando imagens e indicando pares de alinhamento. Uma possível solução seria aplicar um método de pré-alinhamento com malhas chaves sempre que o alinhamento for perdido. Essa medida aumentaria ainda mais a flexibilidade na etapa de captura.
3. **O método proposto como um todo poderia ser adaptado para uma versão paralelizável.** Essa medida daria um ganho expressivo, talvez chegando ao tempo-real. Na implementação atual não é utilizado nenhuma plataforma de computação paralela que aproveite as GPU multicore.
4. **Texturizar o objeto com auxílio de imagens externas ao sensor 3D.** A textura do objeto é uma informação muito importante para preservação digital, infelizmente câmeras RGB-D também são limitadas na qualidade da cor. A implementação de um método que combine textura de câmeras externas de maior qualidade para os modelos já gerados, ofereceria um aspecto muito mais verossímil aos objetos.

Referências Bibliográficas

- [Akca et al., 2007] Akca, D., Grün, A., Breuckmann, B., Lahanier, C., Akca, D., Grün, A., and Grün, A. (2007). *High Definition 3D-scanning of Arts Objects and Paintings*. Institute of Geodesy and Photogrammetry.
- [Alexiadis et al., 2013] Alexiadis, D. S., Zarpalas, D., and Daras, P. (2013). Real-time, full 3-d reconstruction of moving foreground objects from multiple consumer depth cameras. *IEEE Transactions on Multimedia*, 15(2):339–358.
- [Allen et al., 2003] Allen, P. K., Troccoli, A., Smith, B., Stamos, I., and Murray, S. (2003). The beauvais cathedral project. volume 1.
- [Andrade et al., 2012a] Andrade, B. T., Bellon, O. R. P., Silva, L., and Vrubel, A. (2012a). Digital preservation of brazilian indigenous artworks: Generating high quality textures for 3D models. *Journal of Cultural Heritage*, 13(1):28 – 39.
- [Andrade et al., 2012b] Andrade, B. T., Mendes, C. M., Santos Junior, J. O., Bellon, O. R. P., and Silva, L. (2012b). 3D preserving xviii century barroque masterpiece: Challenges and results on the digital preservation of aleijadinho’s sculpture of the prophet joel. *Journal of Cultural Heritage*, 13(2):210 – 214.
- [Arbace et al., 2012] Arbace, L., Elisabetta, S., Callieri, M., Dellepiane, M., Fabbri, M., Antonio, I. I., and Scopigno, R. (2012). Innovative uses of 3D digital technologies to assist the restoration of a fragmented terracotta statue. *Journal of Cultural Heritage*, Online first.
- [Baracchini et al., 2004] Baracchini, C., Brogi, A., Callieri, M., Capitani, L., Cignoni, P., Fasano, A., Montani, C., Nenci, C., Novello, R. P., Pingi, P., Ponchio, F., and Scopigno, R. (2004).

- Digital reconstruction of the Arrigo VII funerary complex. In *Proceedings of Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pages 145–154.
- [Basri and Jacobs, 2001] Basri, R. and Jacobs, D. (2001). Photometric stereo with general, unknown lighting. In *Proceedings of Conference on Computer Vision and Pattern Recognition*.
- [Bathow et al., 2010] Bathow, C., Breuckmann, B., Callieri, M., Corsini, M., Dellepiane, M., Dercks, U., Scopigno, R., and Sigismondi, R. (2010). Documenting and monitoring small fractures on michelangelo’s david. In *Proceeding of the Computer Assisted Assessment*, pages 23–26.
- [Bernardini and Rushmeier, 2002] Bernardini, F. and Rushmeier, H. (2002). The 3D model acquisition pipeline. *Computer Graphics Forum*, 21(2):149–172.
- [Bernardini et al., 2002] Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J., and Taubin, G. (2002). Building a digital model of michelangelo’s florentine pieta. *IEEE Computer Graphics and Applications*, 22(1):59–67.
- [Besl, 1988] Besl, P. J. (1988). *Surfaces in range image understanding*. Springer.
- [Bhavsar and Rajagopalan, 2012] Bhavsar, A. V. and Rajagopalan, A. N. (2012). Range map superresolution-inpainting, and reconstruction from sparse data. *Computer Vision and Image Understanding*, 116(4):572 – 591.
- [Blais, 2004] Blais, F. (2004). Review of 20 years of range sensor development. *J. Electronic Imaging*, 13(1):231–243.
- [Bok et al., 2011] Bok, Y., Jeong, Y., Choi, D., and Kweon, I. (2011). Capturing village-level heritages with a hand-held camera-laser fusion sensor. *International Journal of Computer Vision*, 94(1):36–53.
- [Bradski, 2000] Bradski, G. (2000). Opencv. *Dr. Dobb’s Journal of Software Tools*.
- [Brown et al., 2003] Brown, M. Z., Burschka, D., and Hager, G. D. (2003). Advances in computational stereo. *IEEE Pattern Analysis and Machine Intelligence*, 25(8):993–1008.

- [Callieri et al., 2009] Callieri, M., Cignoni, P., Dellepiane, M., and Scopigno, R. (2009). Pushing time-of-flight scanners to the limit. In *Proceedings of the Conference on Virtual Reality, Archaeology and Cultural Heritage*, pages 85–92.
- [Cappelletto et al., 2013] Cappelletto, E., Zanuttigh, P., and Cortelazzo, G. M. (2013). Handheld scanning with 3d cameras. In *International Workshop on Multimedia Signal Processing*, pages 367–372.
- [Caspi et al., 1998] Caspi, D., Kiryati, N., and Shamir, J. (1998). Range imaging with adaptive color structured light. *IEEE Pattern Analysis and Machine Intelligence*, 20(5):470–480.
- [Chatterjee et al., 2012] Chatterjee, A., Jain, S., and Govindu, V. M. (2012). A pipeline for building 3d models using depth cameras. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, pages 38:1–38:8.
- [Chen et al., 2013] Chen, J., Bautembach, D., and Izadi, S. (2013). Scalable real-time volumetric surface reconstruction. *ACM Transactions on Graphics*, 32(4):113:1–113:16.
- [Chen et al., 2014] Chen, K., Lai, Y., Wu, Y., Martin, R., and Hu, S. (2014). Automatic semantic modeling of indoor scenes from low-quality rgb-d data using contextual information. *ACM Transaction Graphics*, 33(6):208:1–208:12.
- [Chen and Medioni, 1992] Chen, Y. and Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155.
- [Chow and Lichti, 2013] Chow, J. C. K. and Lichti, D. D. (2013). Photogrammetric bundle adjustment with self-calibration of the primesense 3d camera technology: Microsoft kinect. *Access, IEEE*, 1:465–474.
- [Cui et al., 2010] Cui, Y., Schuon, S., Chan, D., Thrun, S., and Theobalt, C. (2010). 3d shape scanning with a time-of-flight camera. In *Conference on Computer Vision and Pattern Recognition*, pages 1173–1180.

- [Cui et al., 2013] Cui, Y., Schuon, S., Thrun, S., Stricker, D., and Theobalt, C. (2013). Algorithms for 3d shape scanning with a depth camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1039–1050.
- [Curless, 2000] Curless, B. (2000). 3D Photography. Technical report, Special Interest Group on Graphics and Interactive Techniques. Course Notes.
- [Dame et al., 2013] Dame, A., Prisacariu, V., Ren, C., and Reid, I. (2013). Dense reconstruction using 3d object shape priors. pages 1288–1295.
- [Fontana et al., 2002] Fontana, R., Greco, M., Materazzi, M., Pampaloni, E., Pezzati, L., Rocchini, C., and Scopigno, R. (2002). Three-dimensional modelling of statues: the minerva of arezzo. *Journal of Cultural Heritage*, 3(4):325 – 331.
- [Frahm et al., 2010] Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S., and Pollefeys, M. (2010). Building Rome on a Cloudless Day. In *Proceedings of European Conference on Computer Vision*, pages 368–381.
- [Godin et al., 2002] Godin, G., Beraldin, J. A., Taylor, J., Cournoyer, L., Rioux, M., El-Hakim, S., Baribeau, R., Blais, F., Boulanger, P., Domey, J., and Picard, M. (2002). Active optical 3d imaging for heritage applications. *IEEE Computer Graphics and Applications*, 22(5):24–36.
- [Gomes et al., 2014] Gomes, L., Bellon, O. R. P., and Silva, L. (2014). 3d reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, (0).
- [Grün et al., 2002] Grün, A., Remondino, F., Zhang, L., and für Geodäsie und Photogrammetrie (Zürich), I. (2002). *Reconstruction of the Great Buddha of Bamiyan, Afghanistan*. ETH, Eidgenössische Technische Hochschule Zürich, Institute of Geodesy and Photogrammetry.
- [Guhring, 2001] Guhring, J. (2001). Dense 3d surface acquisition by structured light using off-the-shelf components. *Proceedings of SPEI on Videometrics and Optical Methods for 3D Shape Measuring*, 4309:220–231.

- [Han et al., 2013] Han, Y., Lee, J., and Kweon, I. (2013). High quality shape from a single rgb-d image under uncalibrated natural illumination. In *International Conference on Computer Vision*, pages 1617–1624.
- [Hertzmann and Seitz, 2003] Hertzmann, A. and Seitz, S. M. (2003). Shape and materials by example: A photometric stereo approach. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pages 533–540.
- [Hertzmann and Seitz, 2005] Hertzmann, A. and Seitz, S. M. (2005). Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Pattern Analysis and Machine Intelligence*, 27(8):1254–1264.
- [Ho and Newman, 2007] Ho, K. L. and Newman, P. (2007). Detecting loop closure with scene sequences. *International Journal of Computer Vision*, 74(3):261–286.
- [Ikeuchi et al., 2007] Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M., and Okamoto, Y. (2007). The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *Journal of Computer Vision*, 75(1):189–208.
- [Intel, 2016] Intel (2016). Intel - [http://http://intelrealsense.bemyapp.com](http://intelrealsense.bemyapp.com).
- [Iwashita et al., 2005] Iwashita, Y., Kurazume, R., Hasegawa, T., and Hara, K. (2005). Fast alignment of 3D geometrical models and 2D color images using 2D distance maps. In *Proceedings of the Conference on 3D Digital Imaging and Modeling*, pages 164–171.
- [Izadi et al., 2011] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. (2011). Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of ACM symposium on User interface software and technology*, pages 559–568.
- [J. et al., 2004] J., S., J., P., and J., B. (2004). Pattern codification strategies in structured light systems. *Pattern Recognition*, 37:827–849.

- [Jiang and Liu, 2013] Jiang, B. and Liu, X. (2013). A divide-and-conquer approach to large scene reconstruction with interactive scene analysis and segmentation. In *International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pages 283–284.
- [Kazhdan et al., 2006] Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. In *Proceedings of the Eurographics symposium on Geometry processing*, pages 61–70.
- [Khoshelham and Elberink, 2012] Khoshelham, K. and Elberink, S. O. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454.
- [Kubacki et al., 2012] Kubacki, D. B., Bui, H. Q., Babacan, S. D., and Do, M. N. (2012). Registration and integration of multiple depth images using signed distance function. In *Proceedings of the international society for optics and photonics*, volume 8296, pages 82960Z–11.
- [Lakdawalla and Hertzmann, 2007] Lakdawalla, A. and Hertzmann, A. (2007). Shape from video: Dense shape, texture, motion and lighting from monocular image streams. In *IEEE Workshop on Photometric Analysis for Computer Vision*.
- [Laurentini, 1994] Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding. *IEEE Pattern Analysis and Machine Intelligence*, 16(2):150–162.
- [Levoy et al., 2000] Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., and Fulk, D. (2000). The Digital Michelangelo Project: 3D Scanning of Large Statues. In *Proceedings of the Conference on Computer Graphics and Interactive Techniques*, pages 131–144.
- [Livieratos, 1992] Livieratos, E. (1992). Empiric, topographic or photogrammetric recording? answers to properly phrased questions. In *Proceedings of the congress in Terrestrial Photogrammetry and Geographic Information Systems for the documentation of the National Cultural Heritage*.

- [Mara et al., 2010] Mara, H., Krömker, S., Jakob, S., and Breuckmann, B. (2010). Gigamesh and gilgamesh: 3D multiscale integral invariant cuneiform character extraction. In *Proceedings of the Conference on Virtual Reality, Archaeology and Cultural Heritage*, pages 131–138.
- [Matsuo et al., 2013] Matsuo, T., Fukushima, N., and Ishibashi, Y. (2013). Weighted joint bilateral filter with slope depth compensation filter for depth map refinement. In *Computational Geometry: Theory and Applications*, pages 300–309.
- [Microsoft, 2014] Microsoft (2014). Kinect - <http://www.xbox.com/en-us/kinect>.
- [Nakazawa et al., 2012] Nakazawa, M., Mitsugami, I., Makihara, Y., Nakajima, H., Habe, H., Yamazoe, H., and Yagi, Y. (2012). Dynamic scene reconstruction using asynchronous multiple kinects. In *International Conference on Pattern Recognition*, pages 469–472.
- [Nashman et al., 1995] Nashman, M., Hong, T. H., and Herman, M. (1995). An integrated vision touch-probe system for dimensional inspection tasks. In *National Institute of Standards and Technology*, pages 3–6.
- [Niebner et al., 2013] Niebner, M., Zollhofer, M., Izadi, S., and Stamminger, M. (2013). Real-time 3d reconstruction at scale using voxel hashing. *Transaction on Graphics*, 32(6):169:1–169:11.
- [Noguchi et al., 2013] Noguchi, S., Watanabe, Y., and Ishikawa, M. (2013). High-resolution surface reconstruction based on multi-level implicit surface from multiple range images. In *International Conference on Image Processing*, pages 2140–2144.
- [Pamplona Segundo et al., 2012] Pamplona Segundo, M., Gomes, L., Bellon, O. R. P., and Silva, L. (2012). Automating 3D reconstruction pipeline by surf-based alignment. In *Proceedings of Conference on Image Processing*, pages 1761–1764.
- [Pavlidis et al., 2007] Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., and Chamzas, C. (2007). Methods for 3D digitization of cultural heritage. *Journal of Cultural Heritage*, 8(1):93–98.

- [Potmesil, 1987] Potmesil, M. (1987). Generating octree models of 3d objects from their silhouettes in a sequence of images. 40(1):1–29.
- [Pradeep et al., 2013] Pradeep, V., Rhemann, C., Izadi, S., Zach, C., Bleyer, M., and Bathiche, S. (2013). Monofusion: Real-time 3d reconstruction of small scenes with a single web camera. In *International Symposium on Mixed and Augmented Reality*, pages 83–88.
- [Prisacariu et al., 2014] Prisacariu, V. A., Kahler, O., Cheng, M. M., Ren, C. Y., Valentin, J., Torr, P. H. S., Reid, I. D., and Murray, D. W. (2014). A Framework for the Volumetric Integration of Depth Images. *ArXiv e-prints*.
- [Pulli, 1999] Pulli, K. (1999). Multiview registration for large data sets. In *Proceedings of the Conference on 3D Digital Imaging and Modeling*, pages 160–168.
- [ReconstructMe, 2014] ReconstructMe (2014). Reconstructme - <http://reconstructme.net/>.
- [Remondino and Rizzi, 2010] Remondino, F. and Rizzi, A. (2010). Reality-based 3D documentation of natural and cultural heritage sites: techniques, problems, and examples. *Applied Geomatics*, 2(3):85–100.
- [Ren et al., 2013a] Ren, C. Y., Prisacariu, V., Murray, D., and Reid, I. (2013a). Star3d: Simultaneous tracking and reconstruction of 3d objects using rgb-d data. In *International Conference on Computer Vision*.
- [Ren et al., 2013b] Ren, X., Fox, D., and Konolige, K. (2013b). Change their perception: Rgb-d for 3-d modeling and recognition. *IEEE Robotics and Automation Magazine*, 20(4):49–59.
- [Richardt et al., 2012] Richardt, C., Stoll, C., Dodgson, N. A., Seidel, H., and Theobalt, C. (2012). Coherent spatio-temporal filtering, upsampling and rendering of RGBZ videos. In *Computer Graphics Forum (Proceedings of Eurographics)*, volume 31.
- [Rocchini et al., 2001] Rocchini, C., Cignoni, P., Montani, C., Pinci, P., and Scopigno, R. (2001). A low cost 3D scanner based on structured light. volume 20, pages 299–308.

- [Rushmeier, 2006] Rushmeier, H. (2006). Eternal egypt: Experiences and research directions. In *International Workshop on Recording, Modeling and Visualization of Cultural Heritage*, pages 22–27.
- [Rusinkiewicz and Levoy, 2001] Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the ICP algorithm. *Proceedings of the Conference on 3D Digital Imaging and Modeling*, pages 145–152.
- [Salvi et al., 2007] Salvi, J., Matabosch, C., Fofi, D., and Forest, J. (2007). A review of recent range image registration methods with accuracy evaluation. *Image Vision Computing*, 25(5):578–596.
- [Santos Junior et al., 2012] Santos Junior, J. O., Vrabel, A., Bellon, O. R. P., and Silva, L. (2012). 3D reconstruction of cultural heritages: Challenges and advances on precise mesh integration. *Computer Vision and Image Understanding*, 116(12):1195–1207.
- [Scharstein and Szeliski, 2003] Scharstein, D. and Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pages 195–202.
- [Schechner and Kiryati, 1998] Schechner, Y. Y. and Kiryati, N. (1998). Depth from defocus vs. stereo: How different really are they? pages 1784–1786.
- [Silva et al., 2013] Silva, J. W., Gomes, L., Agüero, K. A., Silva, L., and Bellon, O. R. P. (2013). Improving 3d reconstruction for low-cost devices. In *International Conference on image processing*.
- [Sinha et al., 2008] Sinha, S. N., Steedly, D., Szeliski, R., Agrawala, M., and Pollefeys, M. (2008). Interactive 3D architectural modeling from unordered photo collections. In *Proceedings of ACM Special Interest Group on Graphics and Interactive Asia*, pages 159–169.
- [Skanect, 2014] Skanect (2014). Skanect - <http://skanect.manctl.com/>.
- [Smisek et al., 2011] Smisek, J., Jancosek, M., and Pajdla, T. (2011). 3d with kinect. In *International Conference on Computer Vision*, pages 1154–1160.

- [Snavely et al., 2006] Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo tourism: exploring photo collections in 3D. *ACM Transaction on Graphics*, 25(3):835–846.
- [Somanath et al., 2013] Somanath, G., Cohen, S., Price, B. L., and Kambhamettu, C. (2013). Stereo+kinect for high resolution stereo correspondences. In *International Conference on 3D Vision*, pages 9–16.
- [Sonnemann et al., 2006] Sonnemann, T., Sauerbier, M., Remondino, F., Schrotter, G., Campana, S., and Forte, M. (2006). Reality-based 3d modeling of the angkorian temples using aerial images. *British Archaeological Reports international series*, 1568:573–579.
- [Steinbruecker et al., 2013] Steinbruecker, F., Kerl, C., Sturm, J., and Cremers, D. (2013). Large-scale multi-resolution surface reconstruction from rgb-d sequences. In *Conference on Computer Vision and Pattern Recognition*.
- [Szeliski and Kang, 1993] Szeliski, R. and Kang, S. B. (1993). Recovering 3d shape and motion from image streams using nonlinear least squares. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pages 752–753.
- [Tam et al., 2013] Tam, G. K. L., Cheng, Z., Lai, Y., Langbein, F. C., Liu, Y., Marshall, D., Martin, R. R., Sun, X., and Rosin, P. L. (2013). Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *Transactions on Visualization and Computer Graphics*, 19(7):1199–1217.
- [Tomasi and Kanade, 1992] Tomasi, C. and Kanade, T. (1992). Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154.
- [Tosovic et al., 2002] Tosovic, S., Sablatnig, R., and Kampe, M. (2002). On combining shape from silhouette and shape from structured light.
- [Vrubel et al., 2009] Vrubel, A., Bellon, O., and Silva, L. (2009). A 3D reconstruction pipeline for digital preservation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2687–2694.

- [Wang et al., 2014] Wang, K., Wang, X., Pan, Z., and Liu, K. (2014). A two-stage framework for 3d facereconstruction from RGBD images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1493–1504.
- [Whelan et al., 2012] Whelan, T., Kaess, M., Fallon, M. F., Johannsson, H., Leonard, J. J., and McDonald, J. B. (2012). Kintinuous: Spatially extended KinectFusion. In *Robotics Science and Systems Workshop on RGB-D: Advanced Reasoning with Depth Cameras*.
- [Whelan et al., 2013] Whelan, T., Kaess, M., Leonard, J. J., and McDonald, J. B. (2013). Deformation-based loop closure for large scale dense RGB-D SLAM. In *International Conference on Intelligent Robots and Systems*.
- [Wu et al., 2014] Wu, C., Zollhofer, M., Niessner, M., Stamminger, M., Izadi, S., and Theobalt, C. (2014). Real-time shading-based refinement for consumer depth cameras. *ACM Transaction Graphics*, 33(6):200:1–200:10.
- [Yi et al., 2012] Yi, W., Jin, W., Hu, J., Tiangang, S., Yang, B., and Zheng, J. (2012). A study in 3d-reconstruction using kinect sensor. In *International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–7.
- [Yilmaz and Karakus, 2013] Yilmaz, O. and Karakus, F. (2013). Stereo and kinect fusion for continuous 3d reconstruction and visual odometry. In *International Conference on Electronics, Computer and Computation*, pages 115–118.
- [Yue et al., 2014] Yue, H., Chen, W., Wu, X., and Liu, J. (2014). Fast 3d modeling in complex environments using a single kinect sensor. *Optics and Lasers in Engineering*, 53:104 – 111.
- [Yutaka and Ohtake, 2003] Yutaka, A. B. and Ohtake, E. B. Y. (2003). A comparison of mesh smoothing methods. In *Conference on Geometric Modeling and Computer Graphics*, pages 83–87.
- [Zhang et al., 2014] Zhang, Q., X., S., Shao, X., Zhao, H., and Shibasaki, R. (2014). When 3d reconstruction meets ubiquitous RGB-D images. In *Computer Vision and Pattern Recognition*, pages 700–707.

- [Zhang et al., 1999] Zhang, R., Tsai, P., Cryer, J. E., and Shah, M. (1999). Shape from shading: A survey. *IEEE Pattern Analysis and Machine Intelligence*, 21(8):690–706.
- [Zhou and Koltun, 2013] Zhou, Q. and Koltun, V. (2013). Dense scene reconstruction with points of interest. *ACM Transactions on Graphics*, 32(4):112:1–112:8.
- [Zollhofer et al., 2014a] Zollhofer, M., Niessner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M., Wu, C., Fitzgibbon, A., Loop, C., Theobalt, C., and Stamminger, M. (2014a). Real-time non-rigid reconstruction using an rgb-d camera. *ACM Transaction Graphics*, 33(4):156:1–156:12.
- [Zollhofer et al., 2014b] Zollhofer, M., Siegl, C., Riffelmacher, B., Vetter, M., Dreyer, B., Stamminger, M., and Bauer, F. (2014b). Low-Cost Real-Time 3D Reconstruction of Large-Scale Excavation Sites using an RGB-D Camera. In *Eurographics Workshop on Graphics and Cultural Heritage*.